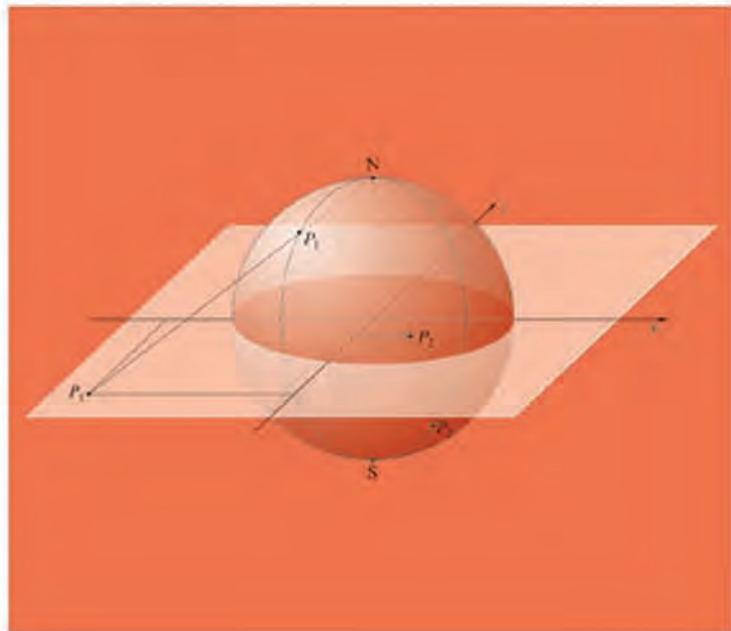Wiley-VCH

Bruce R. Kusse and Erik A. Westwig

Mathematical Physics

Applied Mathematics for Scientists and Engineers
Second Edition

Bruce R. Kusse and Erik A. Westwig

Mathematical Physics

Applied Mathematics for Scientists and Engineers

2nd Edition

WILEY-VCH

This Page Intentionally Left Blank

Bruce R. Kusse and Erik A. Westwig
Mathematical Physics

Related Titles

Vaughn, M. T.
Introduction to Mathematical Physics
2006. Approx. 650 pages with 50 figures.
Softcover
ISBN 3-527-40627-1

Lambourne, R., Tinker, M.
Basic Mathematics for the Physical Sciences
2000. 688 pages.
Softcover
ISBN 0-471-85207-4

Tinker, M., Lambourne, R.
Further Mathematics for the Physical Sciences
2000. 744 pages.
Softcover
ISBN 0-471-86723-3

Courant, R., Hilbert, D.
Methods of Mathematical Physics
Volume 1
1989. 575 pages with 27 figures.
Softcover
ISBN 0-471-50447-5

Volume 2
1989. 852 pages with 61 figures.
Softcover
ISBN 0-471-50439-4

Trigg, G. L. (ed.)
Mathematical Tools for Physicists
2005. 686 pages with 98 figures and 29 tables.
Hardcover
ISBN 3-527-40548-8

Bruce R. Kusse and Erik A. Westwig

Mathematical Physics

Applied Mathematics for Scientists and Engineers

2nd Edition

WILEY-VCH

The Authors

Bruce R. Kusse
College of Engineering
Cornell University
Ithaca, NY
brk2@cornell.edu

Erik Westwig
Palisade Corporation
Ithaca, NY
ewestwig@palisade.com

PREFACE

This book is the result of a sequence of two courses given in the School of Applied and Engineering Physics at Cornell University. The intent of these courses has been to cover a number of intermediate and advanced topics in applied mathematics that are needed by science and engineering majors. The courses were originally designed for junior level undergraduates enrolled in Applied Physics, but over the years they have attracted students from the other engineering departments, as well as physics, chemistry, astronomy and biophysics students. Course enrollment has also expanded to include freshman and sophomores with advanced placement and graduate students whose math background has needed some reinforcement.

While teaching this course, we discovered a gap in the available textbooks we felt appropriate for Applied Physics undergraduates. There are many good introductory calculus books. One such example is *Calculus and Analytic Geometry* by Thomas and Finney, which we consider to be a prerequisite for our book. There are also many good textbooks covering advanced topics in mathematical physics such as *Mathematical Methods for Physicists* by Arfken. Unfortunately, these advanced books are generally aimed at graduate students and do not work well for junior level undergraduates. It appeared that there was no intermediate book which could help the typical student make the transition between these two levels. Our goal was to create a book to fill this need.

The material we cover includes intermediate topics in linear algebra, tensors, curvilinear coordinate systems, complex variables, Fourier series, Fourier and Laplace transforms, differential equations, Dirac delta-functions, and solutions to Laplace's equation. In addition, we introduce the more advanced topics of contravariance and covariance in nonorthogonal systems, multi-valued complex functions described with branch cuts and Riemann sheets, the method of steepest descent, and group theory. These topics are presented in a unique way, with a generous use of illustrations and

graphs and an informal writing style, so that students at the junior level can grasp and understand them. Throughout the text we attempt to strike a healthy balance between mathematical completeness and readability by keeping the number of formal proofs and theorems to a minimum. Applications for solving real, physical problems are stressed. There are many examples throughout the text and exercises for the students at the end of each chapter.

Unlike many text books that cover these topics, we have used an organization that is fundamentally pedagogical. We consider the book to be primarily a teaching tool, although we have attempted to also make it acceptable as a reference. Consistent with this intent, the chapters are arranged as they have been taught in our two course sequence, rather than by topic. Consequently, you will find a chapter on tensors and a chapter on complex variables in the first half of the book and two more chapters, covering more advanced details of these same topics, in the second half. In our first semester course, we cover chapters one through nine, which we consider more important for the early part of the undergraduate curriculum. The last six chapters are taught in the second semester and cover the more advanced material.

We would like to thank the many Cornell students who have taken the AEP 321/322 course sequence for their assistance in finding errors in the text, examples, and exercises. E.A.W. would like to thank Ralph Westwig for his research help and the loan of many useful books. He is also indebted to his wife Karen and their son John for their infinite patience.

BRUCE R. KUSSE
ERIK A. WESTWIG

Ithaca, New York

CONTENTS

This Page Intentionally Left Blank

1

A REVIEW OF VECTOR AND MATRIX ALGEBRA USING SUBSCRIPT/SUMMATION CONVENTIONS

This chapter presents a quick review of vector and matrix algebra. The intent is not to cover these topics completely, but rather use them to introduce subscript notation and the Einstein summation convention. These tools simplify the often complicated manipulations of linear algebra.

1.1 NOTATION

Standard, consistent notation is a very important habit to form in mathematics. Good notation not only facilitates calculations but, like dimensional analysis, helps to catch and correct errors. Thus, we begin by summarizing the notational conventions that will be used throughout this book, as listed in Table 1.1.

TABLE 1.1. Notational Conventions

Symbol	Quantity
a	A real number
$\underline{B}$	A complex number
A_i	A vector component
$M_{i \cdots j}$	A matrix or tensor element
$[M]$	An entire matrix
$\overline{A}$	A vector
$\hat{e}_i$	A basis vector
$\overline{\overline{T}}$	A tensor
$\mathcal{L}$	An operator

A three-dimensional vector $\overline{V}$ can be expressed as

$$\overline{V} = V_x\hat{e}_x + V_y\hat{e}_y + V_z\hat{e}_z, \tag{1.1}$$

where the components (V_x, V_y, V_z) are called the Cartesian components of $\overline{V}$ and $(\hat{e}_x, \hat{e}_y, \hat{e}_z)$ are the basis vectors of the coordinate system. This notation can be made more efficient by using *subscript notation*, which replaces the letters (x, y, z) with the numbers $(1, 2, 3)$. That is, we define:

$$\begin{array}{lll} \hat{e}_1 \equiv \hat{e}_x & & V_1 \equiv V_x \\ \hat{e}_2 \equiv \hat{e}_y & \text{and} & V_2 \equiv V_y \\ \hat{e}_3 \equiv \hat{e}_z & & V_3 \equiv V_z \end{array} \quad . \tag{1.2}$$

Equation 1.1 becomes

$$\overline{V} = V_1\hat{e}_1 + V_2\hat{e}_2 + V_3\hat{e}_3, \tag{1.3}$$

or more succinctly,

$$\overline{V} = \sum_{i=1,2,3} V_i\hat{e}_i. \tag{1.4}$$

Figure 1.1 shows this notational modification on a typical Cartesian coordinate system.

Although subscript notation can be used in many different types of coordinate systems, in this chapter we limit our discussion to Cartesian systems. Cartesian basis vectors are orthonormal and position independent. *Orthonormal* means the magnitude of each basis vector is unity, and they are all perpendicular to one another. *Position independent* means the basis vectors do not change their orientations as we move around in space. Non-Cartesian coordinate systems are covered in detail in Chapter 3.

Equation 1.4 can be compacted even further by introducing the *Einstein summation convention*, which assumes a summation any time a subscript is repeated in the same term. Therefore,

$$\overline{V} = \sum_{i=1,2,3} V_i\hat{e}_i \equiv V_i\hat{e}_i. \tag{1.5}$$
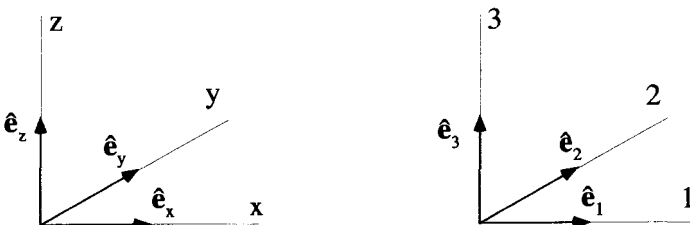
Figure 1.1   The Standard Cartesian System

We refer to this combination of the subscript notation and the summation convention as *subscript/summation notation*.

Now imagine we want to write the simple vector relationship

$$\overline{\mathbf{C}} = \overline{\mathbf{A}} + \overline{\mathbf{B}}. \tag{1.6}$$

This equation is written in what we call *vector notation*. Notice how it does not depend on a choice of coordinate system. In a particular coordinate system, we can write the relationship between these vectors in terms of their components:

$$C_1 = A_1 + B_1$$
$$C_2 = A_2 + B_2 \tag{1.7}$$
$$C_3 = A_3 + B_3.$$

With subscript notation, these three equations can be written in a single line,

$$C_i = A_i + B_i, \tag{1.8}$$

where the subscript i stands for any of the three values $(1, 2, 3)$. As you will see in many examples at the end of this chapter, the use of the subscript/summation notation can drastically simplify the derivation of many physical and mathematical relationships. Results written in subscript/summation notation, however, are tied to a particular coordinate system, and are often difficult to interpret. For these reasons, we will convert our final results back into vector notation whenever possible.

A matrix is a two-dimensional array of quantities that may or may not be associated with a particular coordinate system. Matrices can be expressed using several different types of notation. If we want to discuss a matrix in its entirety, without explicitly specifying all its elements, we write it in *matrix notation* as $[M]$. If we do need to list out the elements of $[M]$, we can write them as a rectangular array inside a pair of brackets:

$$[M] = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1c} \\ M_{21} & M_{22} & \cdots & M_{2c} \\ \vdots & \vdots & \vdots & \vdots \\ M_{r1} & M_{r2} & \cdots & M_{rc} \end{bmatrix}. \tag{1.9}$$

We call this *matrix array notation*. The individual element in the second row and third column of $[M]$ is written as M_{23}. Notice how the row of a given element is always listed first, and the column second. Keep in mind, the array is not necessarily square. This means that for the matrix in Equation 1.9, r does not have to equal c.

Multiplication between two matrices is only possible if the number of columns in the premultiplier equals the number of rows in the postmultiplier. The result of such a multiplication forms another matrix with the same number of rows as the premultiplier and the same number of columns as the postmultiplier. For example, the product between a 3×2 matrix $[M]$ and a 2×3 matrix $[N]$ forms the 3×3 matrix

$[P]$, with the elements given by:

$$
\underbrace{\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \\ M_{31} & M_{32} \end{bmatrix}}_{[M]} \underbrace{\begin{bmatrix} N_{11} & N_{12} & N_{13} \\ N_{21} & N_{22} & N_{23} \end{bmatrix}}_{[N]}
$$

$$
= \underbrace{\begin{bmatrix} M_{11}N_{11} + M_{12}N_{21} & M_{11}N_{12} + M_{12}N_{22} & M_{11}N_{13} + M_{12}N_{23} \\ M_{21}N_{11} + M_{22}N_{21} & M_{21}N_{12} + M_{22}N_{22} & M_{21}N_{13} + M_{22}N_{23} \\ M_{31}N_{11} + M_{32}N_{21} & M_{31}N_{12} + M_{32}N_{22} & M_{31}N_{13} + M_{32}N_{23} \end{bmatrix}}_{[P]} . \quad (1.10)
$$

The multiplication in Equation 1.10 can be written in the abbreviated matrix notation as

$$
[M][N] = [P]. \tag{1.11}
$$

We can also use subscript/summation notation to write the same product as

$$
M_{ij}N_{jk} = P_{ik}, \tag{1.12}
$$

with the implied sum over the j index keeping track of the summation. Notice j is in the second position of the M_{ij} term and the first position of the N_{jk} term, so the summation is over the columns of $[M]$ and the rows of $[N]$, just as it was in Equation 1.10. Equation 1.12 is an expression for the ik^{th} element of the matrix $[P]$.

Matrix array notation is convenient for doing numerical calculations, especially when using a computer. When deriving the relationships between the various quantities in physics, however, matrix notation is often inadequate because it lacks a mechanism for keeping track of the geometry of the coordinate system. For example, in a particular coordinate system, the vector $\overline{V}$ might be written as

$$
\overline{V} = 1\hat{e}_1 + 3\hat{e}_2 + 2\hat{e}_3. \tag{1.13}
$$

When performing calculations, it is sometimes convenient to use a matrix representation of this vector by writing:

$$
\overline{V} \rightarrow [V] = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} . \tag{1.14}
$$

The problem with this notation is that there is no convenient way to incorporate the basis vectors into the matrix. This is why we are careful to use an arrow ($\rightarrow$) in Equation 1.14 instead of an equal sign ($=$). In this text, an equal sign between two quantities means that they are perfectly equivalent in every way. One quantity may be substituted for the other in any expression. For instance, Equation 1.13 implies that the quantity $1\hat{e}_1 + 3\hat{e}_2 + 2\hat{e}_3$ can replace $\overline{V}$ in any mathematical expression, and vice-versa. In contrast, the arrow in Equation 1.14 implies that $[V]$ can represent $\overline{V}$, and that calculations can be performed using it, but we must be careful not to directly substitute one for the other without specifying the basis vectors associated with the components of $[V]$.

1.2 VECTOR OPERATIONS

In this section, we investigate several vector operations. We will use all the different forms of notation discussed in the previous section in order to illustrate their differences. Initially, we will concentrate on matrix and matrix array notation. As we progress, the subscript/summation notation will be used more frequently.

As we discussed earlier, a three-dimensional vector $\overline{V}$ can be represented using a matrix. There are actually two ways to write this matrix. It can be either a (3×1) column matrix or a (1×3) row matrix, whose elements are the components of the vector in some Cartesian basis:

$$\overline{V} \rightarrow [V] = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} \quad \text{or} \quad \overline{V} \rightarrow [V]^\dagger = [\ V_1 \quad V_2 \quad V_3\ ]. \tag{1.15}$$

The standard notation $[V]^\dagger$ has been used to indicate the transpose of $[V]$, indicating an interchange of rows and columns. Remember the vector $\overline{V}$ can have an infinite number of different matrix array representations, each written with respect to a different coordinate basis.

1.2.1 Vector Rotation

Consider the simple rotation of a vector in a Cartesian coordinate system. This example will be worked out, without any real loss of generality, in two dimensions.

We start with the vector $\overline{A}$, which is oriented at an angle θ to the 1-axis, as shown in Figure 1.2. This vector can be written in terms of its Cartesian components as

$$\overline{A} = A_1 \hat{e}_1 + A_2 \hat{e}_2, \tag{1.16}$$

where

$$A_1 = A \cos \theta \qquad A_2 = A \sin \theta. \tag{1.17}$$

In these expressions $A \equiv |\overline{A}| = \sqrt{A_1^2 + A_2^2}$ is the magnitude of the vector $\overline{A}$. The vector $\overline{A}'$ is generated by rotating the vector $\overline{A}$ counterclockwise by an angle ϕ. This
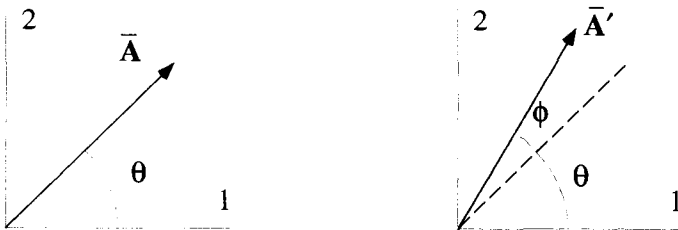
Figure 1.2 Geometry for Vector Rotation

changes the orientation of the vector, but not its magnitude. Therefore, we can write

$$\overline{\mathbf{A}}' = \underbrace{A\cos(\theta + \phi)\hat{\mathbf{e}}_1}_{A_1'} + \underbrace{A\sin(\theta + \phi)\hat{\mathbf{e}}_2}_{A_2'}. \tag{1.18}$$

The components A_1' and A_2' can be rewritten using the trigonometric identities for the cosine and sine of summed angles as

$$A_1' = A\cos(\theta + \phi) = \underbrace{A\cos\theta\cos\phi}_{A_1} - \underbrace{A\sin\theta\sin\phi}_{A_2}$$

$$A_2' = A\sin(\theta + \phi) = \underbrace{A\cos\theta\sin\phi}_{A_1} + \underbrace{A\sin\theta\cos\phi}_{A_2}. \tag{1.19}$$

If we represent $\overline{\mathbf{A}}$ and $\overline{\mathbf{A}}'$ with column matrices,

$$\overline{\mathbf{A}} \rightarrow [A] = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \qquad \overline{\mathbf{A}}' \rightarrow [A'] = \begin{bmatrix} A_1' \\ A_2' \end{bmatrix}, \tag{1.20}$$

Equations 1.19 can be put into matrix array form as

$$\begin{bmatrix} A_1' \\ A_2' \end{bmatrix} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}. \tag{1.21}$$

In the abbreviated matrix notation, we can write this as

$$[A'] = [R(\phi)][A]. \tag{1.22}$$

In this last expression, $[R(\phi)]$ is called the rotation matrix, and is clearly defined as

$$[R(\phi)] = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}. \tag{1.23}$$

Notice that for Equation 1.22 to be the same as Equation 1.19, and for the matrix multiplication to make sense, the matrices $[A]$ and $[A']$ must be column matrix arrays and $[R(\phi)]$ must premultiply $[A]$. The result of Equation 1.19 can also be written using the row representations for $\overline{\mathbf{A}}$ and $\overline{\mathbf{A}}'$. In this case, the transposes of $[R]$, $[A]$ and $[A']$ must be used, and $[R]^\dagger$ must postmultiply $[A]^\dagger$:

$$[A']^\dagger = [A]^\dagger [R]^\dagger. \tag{1.24}$$

Written out using matrix arrays, this expression becomes

$$[\, A_1' \quad A_2' \,] = [\, A_1 \quad A_2 \,] \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}. \tag{1.25}$$

It is easy to see Equation 1.25 is entirely equivalent to Equation 1.21.

These same manipulations can be accomplished using subscript/summation notation. For example, Equation 1.19 can be expressed as

$$A_i' = R_{ij}A_j. \tag{1.26}$$

The matrix multiplication in Equation 1.22 sums over the columns of the elements of $[R]$. This is accomplished in Equation 1.26 by the implied sum over j. Unlike matrix notation, in subscript/summation notation the order of A_j and R_{ij} is no longer important, because

$$R_{ij}A_j = A_j R_{ij}. \tag{1.27}$$

The vector $\overline{\mathbf{A}}'$ can be written using the subscript notation by combining Equation 1.26 with the basis vectors

$$\overline{\mathbf{A}}' = R_{ij}A_j\hat{\mathbf{e}}_i. \tag{1.28}$$

This expression demonstrates a "notational bookkeeping" property of the subscript notation. Summing over a subscript removes its dependence from an expression, much like integrating over a variable. For this reason, the process of subscript summation is often called *contracting* over an index. There are two sums on the right-hand side (RHS) of Equation 1.28, one over the i and another over j. After contraction over both subscripts, the are no subscripts remaining on the RHS. This is consistent with the fact that there are no subscripts on the left-hand side (LHS). The only notation on the LHS is the "overbar" on $\overline{\mathbf{A}}'$, indicating a vector, which also exists on the RHS in the form of a "hat" on the basis vector $\hat{\mathbf{e}}_i$. This sort of notational analysis can be applied to all equations. The notation on the LHS of an equals sign must always agree with the notation on the RHS. This fact can be used to check equations for accuracy. For example,

$$\overline{\mathbf{A}}' \neq R_{ij}A_j, \tag{1.29}$$

because a subscript i remains on the RHS after contracting over j, while there are no subscripts at all on the LHS. In addition, the notation indicates the LHS is a vector quantity, while the RHS is not.

1.2.2 Vector Products

We now consider the dot and cross products of two vectors using subscript/summation notation. These products occur frequently in physics calculations, at every level. The dot product is usually first encountered when calculating the work W done by a force $\overline{\mathbf{F}}$ in the line integral

$$W = \int d\overline{\mathbf{r}} \cdot \overline{\mathbf{F}}. \tag{1.30}$$

In this equation, $d\bar{r}$ is a differential displacement vector. The cross product can be used to find the force on a particle of charge q moving with velocity $\bar{v}$ in an externally applied magnetic field $\bar{B}$:

$$\bar{F} = q(\bar{v} \times \bar{B}). \tag{1.31}$$

The Dot Product   The dot or inner product of two vectors, $\bar{A}$ and $\bar{B}$, is a scalar defined by

$$\bar{A} \cdot \bar{B} = |\bar{A}||\bar{B}| \cos \theta, \tag{1.32}$$

where θ is the angle between the two vectors, as shown in Figure 1.3. If we take the dot product of a vector with itself, we get the magnitude squared of that vector:

$$\bar{A} \cdot \bar{A} = |\bar{A}|^2. \tag{1.33}$$

In subscript/summation notation, Equation 1.32 is written as

$$\bar{A} \cdot \bar{B} = A_i \hat{e}_i \cdot B_j \hat{e}_j. \tag{1.34}$$

Notice we are using two different subscripts to form $\bar{A}$ and $\bar{B}$. This is necessary to keep the summations independent in the manipulations that follow. The notational bookkeeping is working here, because there are no subscripts on the LHS, and none left on the RHS after contraction over both i and j. Only the basis vectors are involved in the dot product, so Equation 1.34 can be rewritten as

$$\bar{A} \cdot \bar{B} = A_i B_j (\hat{e}_i \cdot \hat{e}_j). \tag{1.35}$$

Because we are restricting our attention to Cartesian systems where the basis vectors are orthonormal, we have

$$\hat{e}_i \cdot \hat{e}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}. \tag{1.36}$$

The Kronecker delta,

$$\delta_{ij} \equiv \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}, \tag{1.37}$$
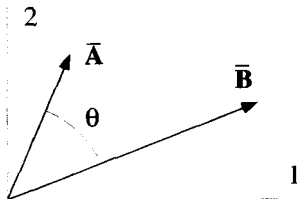
Figure 1.3   The Dot Product

facilitates calculations that involve dot products. Using it, we can write $\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \delta_{ij}$, and Equation 1.35 becomes

$$\overline{\mathbf{A}} \cdot \overline{\mathbf{B}} = A_i B_j \delta_{ij}. \tag{1.38}$$

Equation 1.38 can be expanded by explicitly doing the independent sums over both i and j

$$\overline{\mathbf{A}} \cdot \overline{\mathbf{B}} = A_1 B_1 \delta_{11} + A_1 B_2 \delta_{12} + A_1 B_3 \delta_{13} + A_2 B_1 \delta_{21} + \cdots. \tag{1.39}$$

Since the Kronecker delta is zero unless its subscripts are equal, Equation 1.39 reduces to only three terms,

$$\overline{\mathbf{A}} \cdot \overline{\mathbf{B}} = A_1 B_1 + A_2 B_2 + A_3 B_3 = A_i B_i. \tag{1.40}$$

As one becomes more familiar with the subscript/summation notation and the Kronecker delta, these last steps here are done automatically with the RHS of the brain. Anytime a Kronecker delta exists in a term, with one of its subscripts repeated somewhere else in the same term, the Kronecker delta can be removed, and every instance of the repeated subscript changed to the other subscript of the Kronecker symbol. For example,

$$A_i \delta_{ij} = A_j. \tag{1.41}$$

In Equation 1.38 the Kronecker delta can be grouped with the B_j factor, and contracted over j to give

$$A_i (B_j \delta_{ij}) = A_i B_i. \tag{1.42}$$

Just as well, we could group it with the A_i factor, and sum over i to give an equivalent result:

$$B_j (A_i \delta_{ij}) = B_j A_j. \tag{1.43}$$

This is true for more complicated expressions as well. For example,

$$M_{ij}(A_k \delta_{ik}) = M_{ij} A_i$$

or $$\tag{1.44}$$

$$B_i T_{jk}(\hat{\mathbf{e}}_m \delta_{jm}) = B_i T_{jk} \hat{\mathbf{e}}_j.$$

This flexibility is one of the things that makes calculations performed with subscript/summation notation easier than working with matrix notation.

We should point out that the Kronecker delta can also be viewed as a matrix or matrix array. In three dimensions, this representation becomes

$$\delta_{ij} \to [1] \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{1.45}$$

This matrix can be used to write Equation 1.38 in matrix notation. Notice the contraction over the index i sums over the rows of the [1] matrix, while the contraction

over j sums over the columns. Thus, Equation 1.38 in matrix notation is

$$\overline{A} \cdot \overline{B} \rightarrow [A]^{\dagger}[1][B] = [A_1 \quad A_2 \quad A_3] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}$$

$$= [A]^{\dagger}[B]. \tag{1.46}$$

The Cross Product   The cross or vector product between two vectors $\overline{A}$ and $\overline{B}$ forms a third vector $\overline{C}$, which is written

$$\overline{C} = \overline{A} \times \overline{B}. \tag{1.47}$$

The magnitude of the vector $\overline{C}$ is

$$|\overline{C}| = \|\overline{A}\||\overline{B}| \sin \theta, \tag{1.48}$$

where θ is the angle between the two vectors, as shown in Figure 1.4. The direction of $\overline{C}$ depends on the "handedness" of the coordinate system. By convention, the three-dimensional coordinate systems in physics are usually "right-handed." Extend the fingers of your right hand straight, aligned along the basis vector $\hat{e}_1$. Now, curl your fingers toward the basis vector $\hat{e}_2$. If your thumb now points along $\hat{e}_3$, the coordinate system is right-handed. When the coordinate system is arranged this way, the direction of the cross product follows a similar rule. To determine the direction of $\overline{C}$ in Equation 1.47, point your fingers along $\overline{A}$, and curl them to point along $\overline{B}$. Your thumb is now pointing in the direction of $\overline{C}$. This definition is usually called the *right-hand rule*. Notice, the direction of $\overline{C}$ is always perpendicular to the plane formed by $\overline{A}$ and $\overline{B}$. If, for some reason, we are using a left-handed coordinate system, the definition for the cross product changes, and we must instead use a *left-hand rule*. Because the definition of a cross product changes slightly when we move to
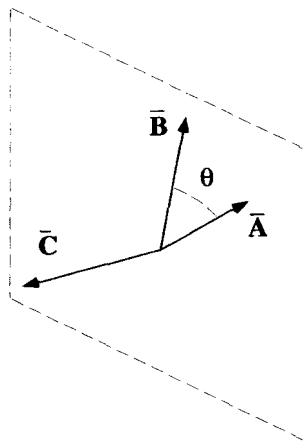
Figure 1.4   The Cross Product

systems of different handedness, the cross product is not exactly a vector, but rather a pseudovector. We will discuss this distinction in more detail at the end of Chapter 4. For now, we will limit our discussion to right-handed coordinate systems, and treat the cross product as an ordinary vector.

Another way to express the cross product is by using an unconventional determinant of a matrix, some of whose elements are basis vectors:

$$\overline{\mathbf{A}} \times \overline{\mathbf{B}} = \begin{vmatrix} \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix}_{det}. \tag{1.49}$$

Expanding the determinant of Equation 1.49 gives

$$\begin{aligned} \overline{\mathbf{A}} \times \overline{\mathbf{B}} = \ & (A_2 B_3 - A_3 B_2)\hat{\mathbf{e}}_1 \\ & + (A_3 B_1 - A_1 B_3)\hat{\mathbf{e}}_2 \\ & + (A_1 B_2 - A_2 B_1)\hat{\mathbf{e}}_3. \end{aligned} \tag{1.50}$$

This last expression can also be written using subscript/summation notation, with the introduction of the Levi-Civita symbol ϵ_{ijk}:

$$\overline{\mathbf{A}} \times \overline{\mathbf{B}} = A_i B_j \hat{\mathbf{e}}_k \epsilon_{ijk}, \tag{1.51}$$

where ϵ_{ijk} is defined as

$$\epsilon_{ijk} = \begin{cases} +1 & \text{for } (i, j, k) = \text{even permutations of } (1, 2, 3) \\ -1 & \text{for } (i, j, k) = \text{odd permutations of } (1, 2, 3) \\ 0 & \text{if two or more of the subscripts are equal} \end{cases} \tag{1.52}$$

An odd permutation of $(1, 2, 3)$ is any rearrangement of the three numbers that can be accomplished with an odd number of pair interchanges. Thus, the odd permutations of $(1, 2, 3)$ are $(2, 1, 3)$, $(1, 3, 2)$, and $(3, 2, 1)$. Similarly, the even permutations of $(1, 2, 3)$ are $(1, 2, 3)$, $(2, 3, 1)$, and $(3, 1, 2)$. Because the subscripts i, j, and k can each independently take the values $(1, 2, 3)$, one way to visualize the Levi-Civita symbol is as the $3 \times 3 \times 3$ array shown in Figure 1.5.
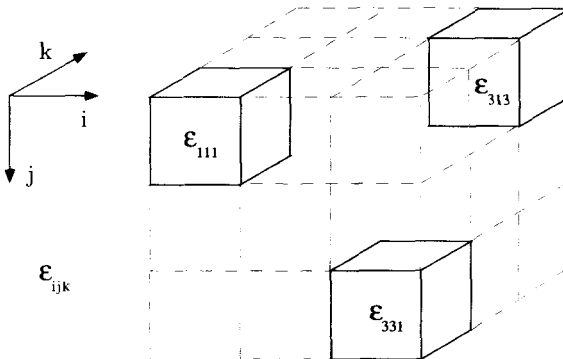
Figure 1.5   The $3 \times 3 \times 3$ Levi-Civita Array

The cross product, written using subscript/summation notation in Equation 1.51, and the dot product, written in the form of Equation 1.38 are very useful for manual calculations, as you will see in the following examples.

1.2.3   Calculations Using Subscript/Summation Notation

We now give two examples to demonstrate the use of subscript/summation notation. The first example shows that a vector's magnitude is unaffected by rotations. The primary function of this example is to show how a derivation performed entirely with matrix notation can also be done using subscript notation. The second derives a common vector identity. This example shows how the subscript notation is a powerful tool for deriving complicated vector relationships.

Example 1.1   Refer back to the rotation picture of Figure 1.2, and consider the products $\overline{A} \cdot \overline{A}$ and $\overline{A}' \cdot \overline{A}'$, first using matrix notation and then using subscript/summation notation. Because $\overline{A}'$ is generated by a simple rotation of $\overline{A}$ we know these two dot products, which represent the magnitude squared of the vectors, should be equal.
   Using matrices:

$$\overline{A} \cdot \overline{A} \rightarrow [A]^\dagger [A] \tag{1.53}$$

$$\overline{A}' \cdot \overline{A}' \rightarrow [A']^\dagger [A']. \tag{1.54}$$

But $[A']$ and $[A']^\dagger$ can be expressed in terms of $[A]$ and $[A]^\dagger$ as

$$[A'] = [R(\phi)][A] \qquad [A']^\dagger = [A]^\dagger [R(\phi)]^\dagger, \tag{1.55}$$

where $R(\phi)$ is the rotation matrix defined in Equation 1.23. If these two equations are substituted into Equation 1.54, we have

$$\overline{A}' \cdot \overline{A}' \rightarrow [A]^\dagger [R(\phi)]^\dagger [R(\phi)][A]. \tag{1.56}$$

The product between the two rotation matrices can be performed,

$$[R(\phi)]^\dagger [R(\phi)] = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{1.57}$$

and Equation 1.56 becomes

$$\overline{A}' \cdot \overline{A}' \rightarrow [A]^\dagger [1][A] = [A]^\dagger [A] \rightarrow \overline{A} \cdot \overline{A}. \tag{1.58}$$

Our final conclusion is that

$$\overline{A}' \cdot \overline{A}' = \overline{A} \cdot \overline{A}. \tag{1.59}$$

To arrive at this result using matrices, care had to be taken to be sure that the matrix operations were in the proper order and that the row, column, and transpose matrices were all used correctly.

Now let's repeat this derivation using the subscript/summation notation. Equation 1.40 allows us to write

$$\overline{\mathbf{A}} \cdot \overline{\mathbf{A}} = A_i A_i \tag{1.60}$$

$$\overline{\mathbf{A}}' \cdot \overline{\mathbf{A}}' = A_r' A_r'. \tag{1.61}$$

Notice how we have been careful to use different subscripts for the two sums in Equations 1.60 and 1.61. This ensures the sums will remain independent as they are manipulated in the following steps. The primed components can be expressed in terms of the unprimed components as

$$A_i' = R_{ij} A_j, \tag{1.62}$$

where R_{ij} is the ij^{th} component of the rotation matrix $R[\phi]$. Inserting this expression into Equation 1.61 gives

$$\overline{\mathbf{A}}' \cdot \overline{\mathbf{A}}' = R_{ru} A_u R_{rv} A_v, \tag{1.63}$$

where again, we have been careful to use the two different subscripts u and v. This equation has three implicit sums, over the subscripts r, u, and v.

In subscript notation, unlike matrix notation, the ordering of the terms is not important, so we rearrange Equation 1.63 to read

$$\overline{\mathbf{A}}' \cdot \overline{\mathbf{A}}' = A_u A_v R_{ru} R_{rv}. \tag{1.64}$$

Next concentrate on the sum over r, which only involves the $[R]$ matrix elements, in the product $R_{ru} R_{rv}$. What exactly does this product mean? Let's compare it to an operation we discussed earlier. In Equation 1.12, we pointed out the subscripted expression $M_{ij} N_{jk}$ represented the regular matrix product $[M][N]$, because the summed subscript j is in the second position of the $[M]$ matrix and the first position of the $[N]$ matrix. The expression $R_{ru} R_{rv}$, however, has a contraction over the first index of *both* matrices. In order to make sense of this product, we write the first instance of $[R]$ using the transpose:

$$R_{ru} R_{rv} \rightarrow [R]^\dagger [R]. \tag{1.65}$$

Consequently, from Equation 1.57,

$$R_{ru} R_{rv} = \delta_{uv}. \tag{1.66}$$

Substituting this result into Equation 1.64 gives

$$\overline{\mathbf{A}}' \cdot \overline{\mathbf{A}}' = A_u A_v \delta_{uv} = A_u A_u = \overline{\mathbf{A}} \cdot \overline{\mathbf{A}}. \tag{1.67}$$

Admittedly, this example is too easy. It does not demonstrate any significant advantage of using the subscript/summation notation over matrices. It does, however, highlight the equivalence of the two approaches. In our next example, the subscript/summation notation will prove to be almost indispensable.

Example 1.2   The subscript/summation notation allows the derivation of vector identities that seem almost impossible using any other approach. The example worked out here is the derivation of an identity for the double cross product between three vectors, $\overline{A} \times (\overline{B} \times \overline{C})$. This one example essentially demonstrates all the common operations that occur in these types of manipulations. Other examples are suggested in the problems listed at the end of this chapter.

The expression $\overline{A} \times (\overline{B} \times \overline{C})$ is written in vector notation and is valid in any coordinate system. To derive our identity, we will convert this expression into subscript/summation notation in a Cartesian coordinate system. In the end, however, we will return our answer to vector notation to obtain a result that does not depend upon any coordinate system. In this example, we will need to use the subscripted form for a vector

$$\overline{V} = V_i \hat{e}_i, \tag{1.68}$$

for a dot product between two vectors

$$\overline{A} \cdot \overline{B} = A_i B_i, \tag{1.69}$$

and for a cross product

$$\overline{A} \times \overline{B} = A_i B_j \hat{e}_k \epsilon_{ijk}. \tag{1.70}$$

To begin, let

$$\overline{D} = \overline{B} \times \overline{C}, \tag{1.71}$$

which, written using the Levi-Civita symbol, is

$$\overline{D} = B_i C_j \hat{e}_k \epsilon_{ijk}. \tag{1.72}$$

Substituting Equation 1.71 into the expression $\overline{A} \times (\overline{B} \times \overline{C})$, and using the Levi-Civita expression again, gives

$$\overline{A} \times (\overline{B} \times \overline{C}) = \overline{A} \times \overline{D} = A_r D_s \hat{e}_t \epsilon_{rst}. \tag{1.73}$$

The sth component of $\overline{D}$ is obtained by dot multiplying both sides of Equation 1.72 by $\hat{e}_s$ as follows:

$$\begin{aligned} D_s = \hat{e}_s \cdot \overline{D} &= \hat{e}_s \cdot B_i C_j \hat{e}_k \epsilon_{ijk} \\ &= B_i C_j \epsilon_{ijk}(\hat{e}_s \cdot \hat{e}_k) \\ &= B_i C_j \epsilon_{ijk} \delta_{sk} \\ &= B_i C_j \epsilon_{ijs}. \end{aligned} \tag{1.74}$$

Substituting the result of Equation 1.74 into Equation 1.73 gives

$$\overline{A} \times (\overline{B} \times \overline{C}) = A_r B_i C_j \epsilon_{ijs} \hat{e}_t \epsilon_{rst}, \tag{1.75}$$

which we rearrange slightly to read

$$\overline{\mathbf{A}} \times (\overline{\mathbf{B}} \times \overline{\mathbf{C}}) = A_r B_i C_j \hat{\mathbf{e}}_t \epsilon_{rst} \epsilon_{ijs}. \tag{1.76}$$

To proceed, some properties of the Levi-Civita symbol need to be developed. First, because of the definition of the Levi-Civita symbol given in Equations 1.52, it is clear that reversing any two of its subscripts just changes its sign, i.e.,

$$\epsilon_{ijk} = -\epsilon_{ikj} = \epsilon_{jki}. \tag{1.77}$$

The second property involves the product of two Levi-Civita symbols that have a common last subscript:

$$\epsilon_{ijk} \epsilon_{mnk} = \delta_{im} \delta_{jn} - \delta_{in} \delta_{jm}. \tag{1.78}$$

With a considerable amount of effort, it can be shown that the RHS of Equation 1.78 has all the properties described by the product of the two Levi-Civita symbols on the LHS, each governed by Equations 1.52. A proof of this identity is given in Appendix A.

With Equations 1.77 and 1.78 we can return to Equation 1.76, which can now be rewritten

$$\overline{\mathbf{A}} \times (\overline{\mathbf{B}} \times \overline{\mathbf{C}}) = A_r B_i C_j \hat{\mathbf{e}}_t (\delta_{rj} \delta_{ti} - \delta_{ri} \delta_{tj}). \tag{1.79}$$

After removing the Kronecker deltas, we obtain

$$\overline{\mathbf{A}} \times (\overline{\mathbf{B}} \times \overline{\mathbf{C}}) = A_j B_i C_j \hat{\mathbf{e}}_i - A_i B_i C_j \hat{\mathbf{e}}_j. \tag{1.80}$$

At this point, you can really see the usefulness of the subscript/summation notation. The factors in the two terms on the RHS of Equation 1.80 can now be rearranged, grouped according to the summations, and returned to vector notation in just two lines! The procedure is:

$$\overline{\mathbf{A}} \times (\overline{\mathbf{B}} \times \overline{\mathbf{C}}) = (A_j C_j)(B_i \hat{\mathbf{e}}_i) - (A_i B_i)(C_j \hat{\mathbf{e}}_j) \tag{1.81}$$

$$= (\overline{\mathbf{A}} \cdot \overline{\mathbf{C}})\overline{\mathbf{B}} - (\overline{\mathbf{A}} \cdot \overline{\mathbf{B}})\overline{\mathbf{C}}. \tag{1.82}$$

Equation 1.81 is valid only in Cartesian systems. But because Equation 1.82 is in vector notation, it is valid in any coordinate system.

In the exercises that follow, you will derive several other vector identities. These will illustrate the power of the subscript/summation notation and help you become more familiar with its use.

EXERCISES FOR CHAPTER 1

1. Consider the two matrices:

$$[M] = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad [N] = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}.$$

With matrix notation, a product between these two matrices can be expressed as $[M][N]$. Using subscript/summation notation, this same product is expressed as $M_{ij}N_{jk}$.

(a) With the elements of the $[M]$ and $[N]$ matrices given above, evaluate the matrix products $[M][N]$, $[N][M]$ and $[M][M]$, leaving the results in matrix array notation.

(b) Express the matrix products of part (a) using the subscript/summation notation.

(c) Convert the following expressions, which are written in subscript/summation notation, to matrix notation:

   i. $M_{jk}N_{ij}$.

   ii. $M_{ij}N_{kj}$.

   iii. $M_{ji}N_{jk}$.

   iv. $M_{ij}N_{jk} + T_{ki}$.

   v. $M_{ji}N_{kj} + T_{ik}$.

2. Consider the square 3×3 matrix $[M]$ whose elements M_{ij} are generated by the expression

$$M_{ij} = ij^2 \quad \text{for } c, j = 1, 2, 3,$$

and a vector $\overline{V}$ whose components in some basis are given by

$$V_k = k \quad \text{for } k = 1, 2, 3.$$

(a) Using a matrix representation for $\overline{V} \rightarrow [V]$, determine the components of the vector that result from a premultiplication of $[V]$ by $[M]$.

(b) Determine the components of the vector that result from a premultiplication of $[M]$ by $[V]^{\dagger}$.

3. The matrix

$$[R] = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

represents a rotation. Show that the matrices $[R]^2 = [R][R]$ and $[R]^3 = [R][R][R]$ are matrices that correspond to rotations of 2θ and 3θ respectively.

4. Let $[D]$ be a 2×2 square matrix and $[V]$ a 2×1 row matrix. Determine the conditions imposed on $[D]$ by the requirement that

$$[D][V] = [V]^{\dagger}[D]$$

for any $[V]$.

5. The trace of a matrix is the sum of all its diagonal elements. Using the subscript/summation notation to represent the elements of the matrices $[T]$ and $[M]$,

(a) Find an expression for the trace of $[T]$.

(b) Show that the trace of the matrix formed by the product $[T][M]$ is equal to the trace of the matrix formed by $[M][T]$.

6. Let $[M]$ be a square matrix. Express the elements of the following matrix products using subscript/summation notation:

(a) $[M][M]^{\dagger}$.

(b) $[M]^{\dagger}[M]$.

(c) $[M]^{\dagger}[M]^{\dagger}$.

7. Convert the following Cartesian, subscript/summation expressions into vector notation:

(a) $V_i A_j B_i \hat{e}_j$.

(b) $c A_i B_j \delta_{ij}$.

(c) $A_l B_j \hat{e}_k \epsilon_{ijk} \delta_{li}$.

(d) $A_i B_j C_l D_m \epsilon_{ijk} \epsilon_{lmk}$.

8. Expand each of the following expressions by explicitly writing out all the terms in the implied summations. Assume (i, j, k) can each take on the values $(1, 2, 3)$.

(a) $\delta_{ij} \epsilon_{ijk}$.

(b) $T_{ij} A_j$.

(c) $\delta_{ij} T_{ij} A_i$.

9. Express the value of the determinant of the matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

using subscript/summation notation and the Levi-Civita symbol.

10. Prove the following vector identities using subscript/summation notation:

(a) $\overline{\mathbf{A}} \cdot (\overline{\mathbf{B}} \times \overline{\mathbf{C}}) = (\overline{\mathbf{A}} \times \overline{\mathbf{B}}) \cdot \overline{\mathbf{C}}$.

(b) $\overline{\mathbf{A}} \times \overline{\mathbf{B}} = -\overline{\mathbf{B}} \times \overline{\mathbf{A}}$.

(c) $(\overline{\mathbf{A}} \times \overline{\mathbf{B}}) \cdot (\overline{\mathbf{C}} \times \overline{\mathbf{D}}) = (\overline{\mathbf{A}} \cdot \overline{\mathbf{C}})(\overline{\mathbf{B}} \cdot \overline{\mathbf{D}}) - (\overline{\mathbf{A}} \cdot \overline{\mathbf{D}})(\overline{\mathbf{B}} \cdot \overline{\mathbf{C}})$.

(d) $(\overline{\mathbf{A}} \times \overline{\mathbf{B}}) \times (\overline{\mathbf{C}} \times \overline{\mathbf{D}}) = [(\overline{\mathbf{A}} \times \overline{\mathbf{B}}) \cdot \overline{\mathbf{D}}]\overline{\mathbf{C}} - [(\overline{\mathbf{A}} \times \overline{\mathbf{B}}) \cdot \overline{\mathbf{C}}]\overline{\mathbf{D}}$.

2

DIFFERENTIAL AND INTEGRAL OPERATIONS ON VECTOR AND SCALAR FIELDS

A field quantity is a function that depends upon space and possibly time. Electric potential, charge density, temperature, and pressure each have only a magnitude, and are described by scalar fields. In contrast, electric and magnetic fields, gravity, current density, and fluid velocity have both magnitude and direction, and are described by vector fields.

Integral and differential operations on vector and scalar field quantities can be expressed concisely using subscript notation and the operator formalism which is introduced in this chapter.

2.1 PLOTTING SCALAR AND VECTOR FIELDS

Because it is frequently useful to picture vector and scalar fields, we begin with a discussion of plotting techniques.

2.1.1 Plotting Scalar Fields

Plots of scalar fields are much easier to construct than plots of vector fields, because scalar fields are characterized by only a single value at each location in time and space. Consider the specific example of the electric potential Φ produced by two parallel, uniform line charges of $\pm\lambda_o$ coulombs per meter that are located at $(x = \pm1, y = 0)$. In meter-kilogram-second (MKS) units, the electric potential produced by this charge distribution is

$$\Phi = \frac{\lambda_o}{4\pi\epsilon_o} \ln\left[\frac{(x + 1)^2 + y^2}{(x - 1)^2 + y^2}\right]. \tag{2.1}$$
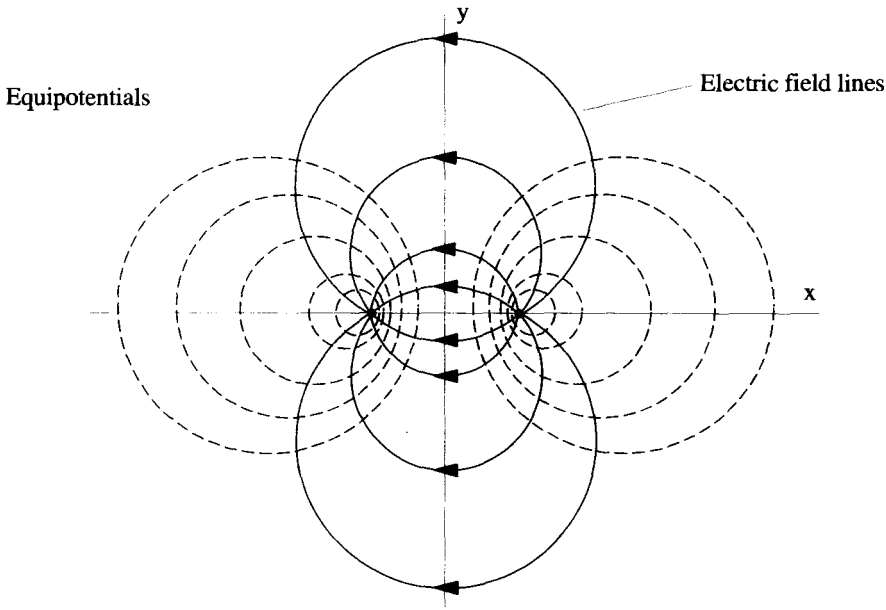
Figure 2.1 Equipotentials and Electric Field Lines of Two Parallel Line Charges

Usually we want to construct surfaces of constant Φ, which in this case are cylinders nested around the line charges. Because there is symmetry in the z direction, these surfaces can be plotted in two dimensions as the dashed, circular contours shown in Figure 2.1. The centers of these circles are located along the x-axis from $1 < x < \infty$ for positive values of Φ, and from $-\infty < x < -1$ for negative values of Φ. $\Phi = 0$ lies on the y-axis, which you can think of as a circle of infinite radius with its center at infinity. If the contours have evenly spaced constant values of Φ, the regions of highest line density show where the function is most rapidly varying with position.

2.1.2 Plotting Vector Fields

Because vectors have both magnitude and direction, plots for their fields are usually much more complicated than for scalar fields. For example, the Cartesian components of the electric field of the preceding example can be calculated to be

$$E_x = -\frac{\partial \Phi}{\partial x} = \frac{\lambda_o}{\pi \epsilon_o} \left[\frac{x^2 - y^2 - 1}{[(x-1)^2 + y^2][(x+1)^2 + y^2]} \right] \qquad (2.2)$$

$$E_y = -\frac{\partial \Phi}{\partial y} = \frac{\lambda_o}{\pi \epsilon_o} \left[\frac{2xy}{[(x-1)^2 + y^2][(x+1)^2 + y^2]} \right]. \qquad (2.3)$$

A vector field is typically drawn by constructing lines which are tangent to the field vector at every point in space. By convention, the density of these field lines

indicates the magnitude of the field, while arrows show its direction. If we suppose an electric field line for Equations 2.2 and 2.3 is given by the equation $y = y(x)$, then

$$\frac{dy(x)}{dx} = \frac{E_y}{E_x} = \frac{2xy}{x^2 - y^2 - 1}.$$ (2.4)

With some work, Equation 2.4 can be integrated to give

$$x^2 + (y - c)^2 = 1 + c^2,$$ (2.5)

where c is the constant of integration. This constant can be varied between $-\infty$ and ∞ to generate the entire family of field lines. For this case, these lines are circles centered on the y-axis at $y = c$ with radii given by $\sqrt{1 + c^2}$. They are shown as the solid lines in Figure 2.1. The arrows indicate how the field points from the positive to the negative charge. Notice the lines are most densely packed directly between the charges where the electric field is strongest.

2.2   INTEGRAL OPERATORS

2.2.1   Integral Operator Notation

The gradient, divergence, and curl operations, which we will review later in this chapter, are naturally in *operator* form. That is, they can be represented by a symbol that operates on another quantity. For example, the gradient of Φ is written as $\overline{\nabla}\Phi$. Here the operator is $\overline{\nabla}$, which acts on the operand Φ to give us the gradient.

In contrast, integral operations are commonly not written in operator form. The integral of $f(x)$ over x is often expressed as

$$\int f(x)\, dx,$$ (2.6)

which is not in operator form because the integral and the operand $f(x)$ are intermingled. We can, however, put Equation 2.6 in operator form by reorganizing the terms in this equation:

$$\int dx\, f(x).$$ (2.7)

Now the operator $\int dx$ acts on $f(x)$ to form the integral, just as the $\overline{\nabla}$ operator acts on Φ to form the gradient. In practice, the integral operator is moved to the right, passing through all the terms of the integrand that do not depend on the integration variable. For example,

$$\int dx\, x^2(x + y)y^2 = y^2 \int dx\, x^2(x + y).$$ (2.8)

2.2.2 Line Integrals

The process of taking an integral along a path is called line integration and is a common operation in all branches of physics. For example, the work that a force $\overline{\mathbf{F}}$ performs as it moves along path C is

$$W = \int_C d\overline{\mathbf{r}} \cdot \overline{\mathbf{F}}. \tag{2.9}$$

Here the line integral operator $\int_C d\overline{\mathbf{r}}$ acts on the force $\overline{\mathbf{F}}$. The differential displacement vector $d\overline{\mathbf{r}}$ is tangent to each point along C, as shown in Figure 2.2. If C closes on itself, we write the operator with a circle on the integral sign:

$$\oint_C d\overline{\mathbf{r}}. \tag{2.10}$$

Because Equation 2.9 is written in vector notation, it is valid in any coordinate system. In a Cartesian system, $d\overline{\mathbf{r}} = dx_i \hat{\mathbf{e}}_i$ and Equation 2.9 becomes

$$W = \int_C d\overline{\mathbf{r}} \cdot \overline{\mathbf{F}} = \int_C dx_i\, F_i. \tag{2.11}$$

Notice that the quantity produced by this integration is a scalar, and the subscript i is summed implicitly.

There are other less common line integral operations. For example, the operator in

$$\int_C d\overline{\mathbf{r}}\, \Phi = \hat{\mathbf{e}}_i \int_C dx_i\, \Phi \tag{2.12}$$

acts on the scalar Φ to produce a vector. Another example,

$$\int_C d\overline{\mathbf{r}} \times \overline{\mathbf{V}} = \hat{\mathbf{e}}_k \int_C dx_i\, V_j \epsilon_{ijk}, \tag{2.13}$$

generates a vector using the cross product. Note that all these subscripted expressions are written in a Cartesian system where the basis vectors are orthonormal and position independent.
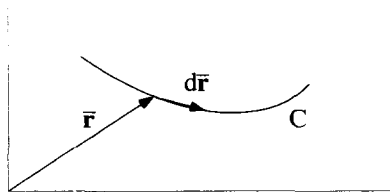
Figure 2.2 The Line Integral

2.2.3 Surface Integrals

Surface integrals involve the integral operator,

$$\int_S d\overline{\sigma}, \tag{2.14}$$

where $d\overline{\sigma}$ is a vector that represents a differential area. This vector has a magnitude equal to a differential area of S, and a direction perpendicular to the surface. If we write the differential area as $d\sigma$ and the unit normal $\hat{n}$, the differential area vector becomes $d\overline{\sigma} = \hat{n}\,d\sigma$. Because a surface has two sides, there is ambiguity in how to define $\hat{n}$. For a simple, closed surface such as the one in Figure 2.3(a), we define $\hat{n}$ to always point in the "outward" direction. If the surface is not closed, i.e., does not enclose a volume, the direction of $\hat{n}$ is typically determined by the closed path C that defines the boundaries of the surface, and the right-hand rule, as shown in Figure 2.3(b).

Frequently, the surface integral operator acts on a vector field quantity by means of the dot product

$$\int_S d\overline{\sigma} \cdot \overline{V}. \tag{2.15}$$

In Cartesian coordinates,

$$d\overline{\sigma} = d\sigma_i \hat{e}_i, \tag{2.16}$$

where $d\sigma_i$ is positive or negative depending on the sign of $\hat{n} \cdot \hat{e}_i$, as discussed in the previous paragraph. This surface integral becomes

$$\int_S d\overline{\sigma} \cdot \overline{V} = \int_S d\sigma_i\, V_i. \tag{2.17}$$

There are less common surface integrals of the form

$$\int_S d\overline{\sigma}\, \Phi = \hat{e}_i \int_S d\sigma_i\, \Phi, \tag{2.18}$$
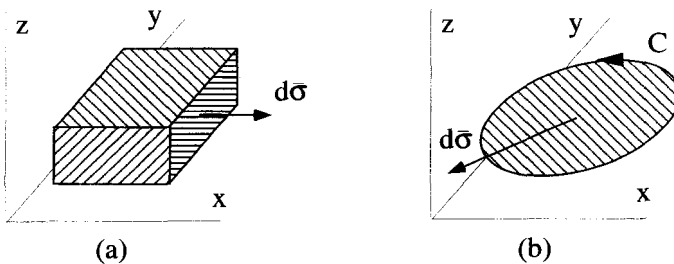
Figure 2.3 Surface Integrals

which is an operation on a scalar which produces a vector quantity, and

$$\int_S d\overline{\boldsymbol{\sigma}} \times \overline{\mathbf{V}} = \hat{\mathbf{e}}_k \int_S d\sigma_i \, V_j \epsilon_{ijk}, \tag{2.19}$$

which also generates a vector.

2.2.4 Volume Integrals

The volume integral is the simplest integral operator because the variables of integration are scalars. It is written

$$\int_V d\tau, \tag{2.20}$$

where $d\tau$ is a differential volume, and V represents the total volume of integration. The most common volume integral acts on a scalar field quantity and, as a result, produces a scalar

$$\int_V d\tau \, \Phi. \tag{2.21}$$

In Cartesian coordinates, this is written

$$\int_V dx_1 \, dx_2 \, dx_3 \, \Phi. \tag{2.22}$$

Volume integrals of vector quantities are also possible:

$$\int_V d\tau \, \overline{\mathbf{V}} = \int_V dx_1 \, dx_2 \, dx_3 \, \overline{\mathbf{V}}. \tag{2.23}$$

2.3 DIFFERENTIAL OPERATIONS

By their definition, field quantities are functions of position. Analogous to how the change in a function of a single variable is described by its derivative, the position dependence of a scalar field can be described by its *gradient*, and the position dependence of a vector field by its *curl* and *divergence*. The Del operator $\overline{\nabla}$ is used to describe all three of these fundamental operations.

The operator $\overline{\nabla}$ is written in coordinate-independent vector notation. It can be expressed in subscript/summation notation in a Cartesian coordinate system as

$$\overline{\nabla} = \hat{\mathbf{e}}_i \frac{\partial}{\partial x_i}. \tag{2.24}$$

Keep in mind, this expression is only valid in Cartesian systems. It will need to be modified when we discuss non-Cartesian coordinate systems in Chapter 3.

When operating on a scalar field, the $\overline{\nabla}$ operator produces a vector quantity called the gradient:

$$\overline{\nabla}\Phi(x_1, x_2, x_3) = \hat{e}_i \frac{\partial \Phi(x_1, x_2, x_3)}{\partial x_i}. \tag{2.25}$$

For example, in electromagnetism the electric field is equal to the negative gradient of the electric potential:

$$\overline{E} = -\overline{\nabla}\Phi = -\hat{e}_i \frac{\partial \Phi(x_1, x_2, x_3)}{\partial x_i}. \tag{2.26}$$

The Del operator also acts upon vector fields via either the dot product or the cross product. The divergence of a vector field is a scalar quantity created using the dot product:

$$\overline{\nabla} \cdot \overline{A} = \hat{e}_i \frac{\partial}{\partial x_i} \cdot A_j \hat{e}_j = \frac{\partial A_i}{\partial x_i}. \tag{2.27}$$

The charge density ρ_c in a region of space can be calculated using the divergence from the relation

$$\rho_c = \epsilon_o (\overline{\nabla} \cdot \overline{E}). \tag{2.28}$$

If instead we used the cross product, we produce a vector quantity called the curl,

$$\overline{\nabla} \times \overline{A} = \left(\hat{e}_i \frac{\partial}{\partial x_i} \right) \times A_j \hat{e}_j = \left(\frac{\partial}{\partial x_i} A_j \right) \hat{e}_k \epsilon_{ijk}, \tag{2.29}$$

where we have used the Levi-Civita symbol to express the cross product in subscript/summation notation. One of Maxwell's equations relates the electric field to the rate of change of the magnetic field using the curl operator:

$$\overline{\nabla} \times \overline{E} = -\frac{\partial \overline{B}}{\partial t}. \tag{2.30}$$

2.3.1 Physical Picture of the Gradient

The gradient of a scalar field is a vector that describes, at each point, how the field changes with position. Dot multiplying both sides of Equation 2.25 by $d\overline{r} = dx_i \hat{e}_i$ gives

$$d\overline{r} \cdot \overline{\nabla}\Phi = dx_i \hat{e}_i \cdot \hat{e}_j \frac{\partial \Phi}{\partial x_j}. \tag{2.31}$$

Manipulating the RHS of this expression gives

$$d\overline{r} \cdot \overline{\nabla}\Phi = \frac{\partial \Phi}{\partial x_i} dx_i. \tag{2.32}$$

The RHS of this expression can be recognized as the total differential change of Φ due to a differential change of position $d\bar{r}$. The result can be written entirely in vector notation as

$$d\Phi = \overline{\nabla}\Phi \cdot d\bar{r}. \tag{2.33}$$

From Equation 2.33, it is clear that the largest value of $d\Phi$ occurs when $d\bar{r}$ points in the same direction as $\overline{\nabla}\Phi$. On the other hand, a displacement perpendicular to $\overline{\nabla}\Phi$ produces no change in Φ because $d\Phi = 0$. This means that the gradient will always point perpendicular to the surfaces of constant Φ.

At the beginning of this chapter, we discussed the electric potential function generated by two line charges. The electric field was generated by taking the gradient of this scalar potential, and was plotted in Figure 2.1. While we could use this example as a model for developing a physical picture for the gradient operation, it is relatively complicated. Instead, we will look at the simpler two-dimensional function

$$\Phi = -xy. \tag{2.34}$$

A plot of the equipotential lines of Φ in the xy-plane is shown in Figure 2.4. The gradient operating on this function generates the vector field

$$\overline{\nabla}\Phi = -y\hat{e}_x - x\hat{e}_y. \tag{2.35}$$

Now imagine we are at the point $(1, 2)$ and move to the right, by an infinitesimal amount dr along the positive x-axis. The corresponding change in Φ can be determined by calculating:

$$\begin{aligned} d\Phi &= \overline{\nabla}\Phi \cdot d\bar{r} \\ &= (-2\hat{e}_x - 1\hat{e}_y) \cdot (dr\,\hat{e}_x) \\ &= -2\,dr. \end{aligned} \tag{2.36}$$
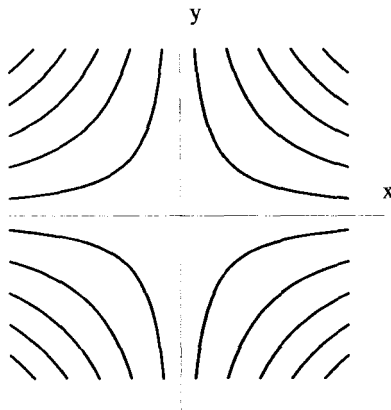
Figure 2.4   Surfaces of Constant $\Phi = -xy$

This says Φ decreases by 2 units for every unit moved in that direction. If instead, we were sitting at the point $(3, 4)$ and move an infinitesimal amount dr, at a 45-degree angle with respect to the x-axis, Φ changes by an amount

$$d\Phi = \overline{\nabla}\Phi \cdot d\overline{r}$$

$$= (-4\hat{e}_x - 3\hat{e}_y) \cdot \frac{dr}{\sqrt{2}}(\hat{e}_x + \hat{e}_y) \qquad (2.37)$$

$$= \frac{-7}{\sqrt{2}} dr.$$

Notice, these changes are for infinitesimal quantities. To calculate the change of Φ over a finite path, where the gradient changes as we move from point to point, we need to use a line integral

$$\Delta\Phi = \int_C d\overline{r} \cdot \overline{\nabla}\Phi. \qquad (2.38)$$

When using the gradient to generate a vector field, we usually add a negative sign to the definition. For example, the electric field is generated from the electrostatic potential by

$$\overline{E} = -\overline{\nabla}\Phi. \qquad (2.39)$$

Using this convention, moving against the field lines increases Φ. For the potential of Equation 2.34, the negative gradient is

$$-\overline{\nabla}\Phi = y\hat{e}_x + x\hat{e}_y. \qquad (2.40)$$

The field lines for this function can be determined as follows:

$$\frac{dy}{dx} = \frac{x}{y}$$

$$dy\, y = dx\, x$$

$$y^2 = x^2 + c$$

$$x^2 - y^2 = c. \qquad (2.41)$$

These lines are perpendicular to the lines of constant Φ, as shown in Figure 2.5. Notice how the density of the vector field lines shows that the magnitude of the field increases as you move away from the origin.

In summary, the gradient of a scalar function Φ generates a vector field which, at each point, indicates the direction of the greatest increase of Φ and lies perpendicular to the lines or surfaces of constant Φ. The example discussed above was worked out in two dimensions, but the process can also be visualized in three dimensions,
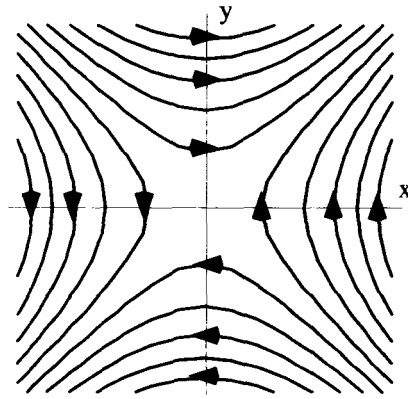
Figure 2.5 Field Lines for $\Phi = -xy$

where $\Phi = constant$ generates surfaces, and the gradient is always normal to these surfaces. Except for visualization, there are no limits to the number of dimensions for the gradient operation.

2.3.2 Physical Picture of the Divergence

The divergence operation will be described physically by developing the *continuity equation*, which describes the change in the local particle density of a fluid as a function of time. Let $\rho(x, y, z, t)$ be the number of particles per unit volume and $\overline{v}(x, y, z, t)$ be the velocity of these particles, at the point (x, y, z) and the time t. Consider a differential volume $d\tau = dx\,dy\,dz$ located at (x_o, y_o, z_o), as shown in Figure 2.6. The continuity equation is obtained by making the assumption that particles can be neither created nor destroyed in this volume, and then equating any net flow of particles into or out of the volume with a corresponding change in ρ. If we call the total number of particles in the infinitesimal volume $N \equiv \rho d\tau$, then

$$\frac{\partial N}{\partial t} = \frac{\partial \rho(x_o, y_o, z_o, t)}{\partial t} \, dx\,dy\,dz. \tag{2.42}$$

The rate of change of N in Equation 2.42 needs to be accounted for by the flow of particles across the six surfaces of $d\tau$. First, consider the bottom cross-hatched surface of Figure 2.6. The flow across this surface can be determined with the aid of Figure 2.7(a). The total number of particles that enter the volume in a time dt across this surface is equal to the number of particles in the shaded region $dx\,dy\,v_z dt$. Notice, a positive v_z adds particles to the volume, while a negative v_z removes them. Thus, the contribution of the bottom surface to $\partial N/\partial t$ is

$$\frac{\partial N_{bottom}}{\partial t} = \rho(x_o, y_o, z_o, t)v_z(x_o, y_o, z_o, t) \, dx\,dy. \tag{2.43}$$
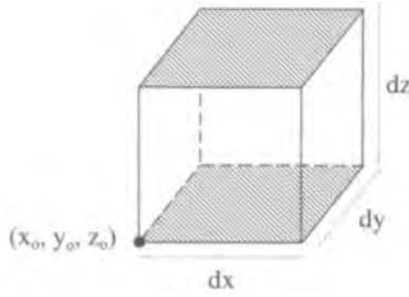
Figure 2.6   Differential Volume

Notice, both ρ and v_z are evaluated at (x_o, y_o, z_o) in this expression. By defining the vector $\bar{J} \equiv \rho \bar{v}$, which is called the current density, Equation 2.43 can be written more concisely as

$$\frac{\partial N_{bottom}}{\partial t} = J_z(x_o, y_o, z_o, t) \, dx \, dy. \tag{2.44}$$

This same type of calculation can be made for the top cross-hatched surface shown in Figure 2.6. Figure 2.7(b) shows that now a positive v_z carries the number of particles in the shaded region out of the volume. This side contributes

$$\frac{\partial N_{top}}{\partial t} = -J_z(x_o, y_o, z_o + dz, t) \, dx \, dy \tag{2.45}$$

to the total $\partial N / \partial t$. Notice, in this case, we evaluate J_z at the point $(x_o, y_o, z_o + dz)$. Combining Equations 2.44 and 2.45 gives

$$\frac{\partial N_{bottom}}{\partial t} + \frac{\partial N_{top}}{\partial t} = \left[J_z(x_o, y_o, z_o, t) - J_z(x_o, y_o, z_o + dz, t) \right] \, dx \, dy. \tag{2.46}$$

(a)   (b)
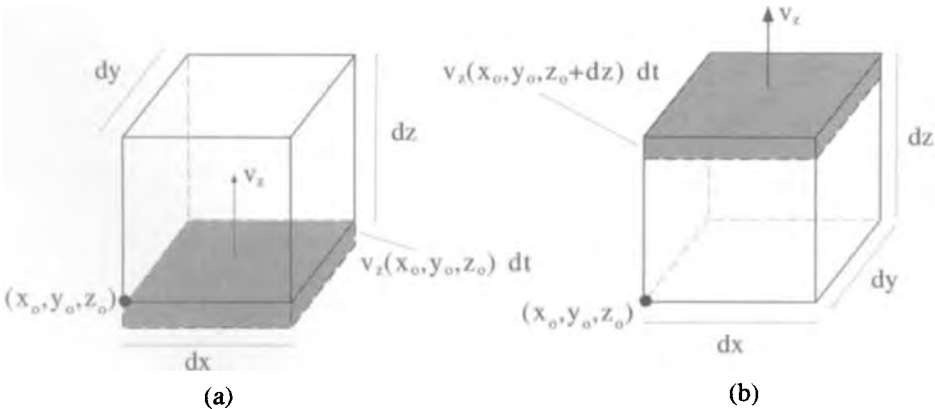
Figure 2.7   Flow Across the Bottom and Top Surfaces

This last expression can be written in terms of the derivative of J_z, because in the differential limit we have

$$J_z(x_o, y_o, z_o + dz, t) = J_z(x_o, y_o, z_o, t) + \left.\frac{\partial J_z}{\partial z}\right|_{(x_o,y_o,z_o)} dz. \qquad (2.47)$$

Substitution of Equation 2.47 into Equation 2.46 gives

$$\frac{\partial N_{bottom}}{\partial t} + \frac{\partial N_{top}}{\partial t} = -\left.\frac{\partial J_z}{\partial z}\right|_{(x_o,y_o,z_o)} dx\, dy\, dz. \qquad (2.48)$$

Working on the other four surfaces produces similar results. The total flow into the differential volume is therefore

$$\frac{\partial N}{\partial t} = \left[-\left.\frac{\partial J_x}{\partial x}\right|_{(x_o,y_o,z_o)} - \left.\frac{\partial J_y}{\partial y}\right|_{(x_o,y_o,z_o)} - \left.\frac{\partial J_z}{\partial z}\right|_{(x_o,y_o,z_o)} \right] dx\, dy\, dz, \qquad (2.49)$$

which can be recognized as $-\overline{\nabla} \cdot \overline{J}$ times $d\tau$. Combining this result with Equation 2.42 gives us the continuity equation:

$$\frac{\partial \rho}{\partial t} = -\overline{\nabla} \cdot \overline{J}. \qquad (2.50)$$

For a positive divergence of $\overline{J}$, more particles are leaving than entering a region so $\partial \rho / \partial t$ is negative.

This exercise provides us with a physical interpretation of the divergence. If the divergence of a vector field is positive in a region, this region is a source. Field lines are "born" in source regions. On the other hand, if the divergence in a region is negative, the region is a sink. Field lines "terminate" in sink regions. If the divergence of a vector field is zero in a region, then all field lines that enter it must also leave.

2.3.3   Physical Picture of the Curl

The curl of a vector field is a vector, which describes *on a local scale* the field's circulation. From the word *curl*, itself, it seems reasonable to conclude that if a vector field has a nonzero curl the field lines should be "curved," while fields with zero curl should be "straight." This is a common misconception. It is possible for vector field lines to appear as shown in Figure 2.8(a), clearly describing a "curved" situation, and have no curl. Also, the fields lines shown in Figure 2.8(b), which are definitely "straight," can have a nonzero curl. To resolve this confusion, we must look at the curl on a differential scale.

Consider a vector field $\overline{V}$ that is a function of only x and y. The curl of this field points in the z-direction, and according to Equation 2.29 is given by

$$\overline{\nabla} \times \overline{V} = \left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} \right)\hat{e}_z \qquad (2.51)$$

for a Cartesian coordinate system.
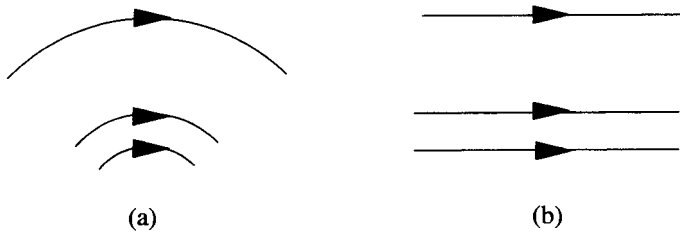
Figure 2.8   Circulating and Noncirculating Vector Fields

Now consider the line integral of a vector field $\overline{V}$ around the closed path shown in Figure 2.9,

$$\oint_C d\overline{r} \cdot \overline{V}. \tag{2.52}$$

The starting point for the integration is the point (x_o, y_o). In this derivation, we need to be a little more careful with the treatment of infinitesimal quantities than we were in the discussion of the divergence. For this reason, we begin by letting the dimensions of the loop be Δx and Δy, as shown in the figure. Later in the derivation, we will shrink these to infinitesimal quantities to obtain our final result.

The integral along C can be broken up into four parts. First consider the integration along C_1, where $y = y_o$ and x varies from x_o to $x_o + \Delta x$:

$$\int_{C_1} d\overline{r} \cdot \overline{V} = \int_{x_o}^{x_o + \Delta x} dx \, V_x. \tag{2.53}$$

Along this segment, we can expand $V_x(x, y_o)$ with a Taylor series, keeping up to the linear term in x:

$$V_x(x, y_o) \approx V_x(x_o, y_o) + \left. \frac{\partial V_x}{\partial x} \right|_{(x_o, y_o)} (x - x_o). \tag{2.54}$$
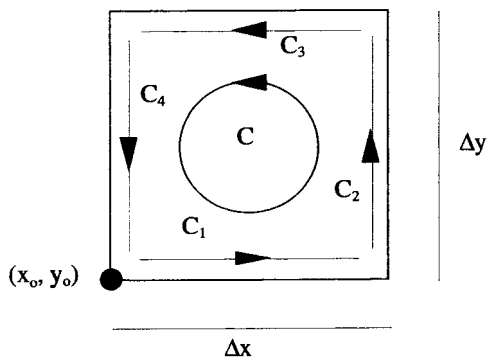
Figure 2.9   Closed Path for Curl Integration

Keeping higher-order terms in this expansion will not change the results of this derivation. Substituting Equation 2.54 into Equation 2.53 and performing the integration gives

$$\int_{C_1} d\bar{\mathbf{r}} \cdot \overline{\mathbf{V}} \approx V_x(x_o, y_o)\Delta x + \frac{1}{2}\frac{\partial V_x}{\partial x}\bigg|_{(x_o,y_o)} (\Delta x)^2. \tag{2.55}$$

Next the integration along C_3, the top section of the loop, is performed. Along this path, y is held fixed at $y = y_o + \Delta y$, while x varies from $x_o + \Delta x$ to x_o. Therefore,

$$\int_{C_3} d\bar{\mathbf{r}} \cdot \overline{\mathbf{V}} = \int_{x_o+\Delta x}^{x_o} dx\, V_x. \tag{2.56}$$

Again, we expand $V_x(x, y_o + \Delta y)$ to first order with a Taylor series:

$$V_x(x, y_o + \Delta y) \approx V_x(x_o, y_o) + \frac{\partial V_x}{\partial x}\bigg|_{(x_o,y_o)} (x - x_o) + \frac{\partial V_x}{\partial y}\bigg|_{(x_o,y_o)} \Delta y. \tag{2.57}$$

Substituting Equation 2.57 into Equation 2.56 and performing the integration gives

$$\int_{C_3} d\bar{\mathbf{r}} \cdot \overline{\mathbf{V}} \approx -V_x(x_o, y_o)\Delta x - \frac{1}{2}\frac{\partial V_x}{\partial x}\bigg|_{(x_o,y_o)} (\Delta x)^2 - \frac{\partial V_x}{\partial y}\bigg|_{(x_o,y_o)} \Delta x\, \Delta y. \tag{2.58}$$

Combining Equations 2.55 and 2.58 gives

$$\int_{C_1} d\bar{\mathbf{r}} \cdot \overline{\mathbf{V}} + \int_{C_3} d\bar{\mathbf{r}} \cdot \overline{\mathbf{V}} \approx -\frac{\partial V_x}{\partial y}\bigg|_{(x_o,y_o)} \Delta x\, \Delta y. \tag{2.59}$$

After performing similar integrations along the C_2 and C_4 paths, we can combine all the results to obtain

$$\oint_C d\bar{\mathbf{r}} \cdot \overline{\mathbf{V}} \approx \left(\frac{\partial V_y}{\partial x}\bigg|_{(x_o,y_o)} - \frac{\partial V_x}{\partial y}\bigg|_{(x_o,y_o)}\right) \Delta x\, \Delta y. \tag{2.60}$$

The error of Equation 2.60 vanishes as the dimensions of the loop shrink to the infinitesimal, $\Delta x \to 0$ and $\Delta y \to 0$. Also, using Equation 2.51, the term in the brackets on the RHS of Equation 2.60 can be identified as the z component of $\overline{\nabla} \times \overline{\mathbf{V}}$. Therefore, we can write

$$\lim_{C \to 0} \oint_C d\bar{\mathbf{r}} \cdot \overline{\mathbf{V}} = \hat{\mathbf{e}}_z \cdot \left(\overline{\nabla} \times \overline{\mathbf{V}}\right) \lim_{S \to 0} \int_S d\sigma_z, \tag{2.61}$$

where C is the contour that encloses S and $d\sigma_z \equiv dx\, dy$ is a differential area of that surface.

What does all this tell us about the curl? The result in Equation 2.61 can be reorganized as:

$$\hat{\mathbf{e}}_z \cdot \left(\overline{\nabla} \times \overline{\mathbf{V}}\right) = \lim_{C,S \to 0} \frac{\oint_C d\bar{\mathbf{r}} \cdot \overline{\mathbf{V}}}{\int_S d\sigma_z}. \tag{2.62}$$

(a)   (b)
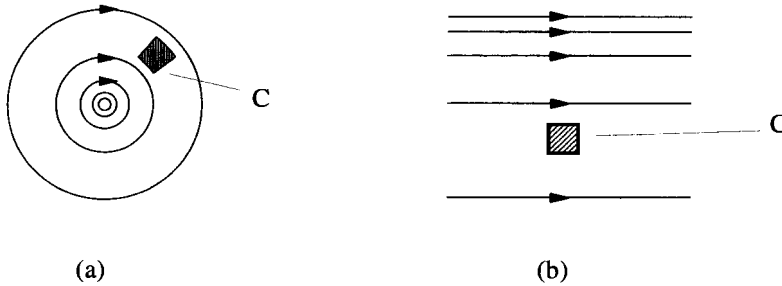
Figure 2.10   Fields with Zero (a) and Nonzero (b) Curl

This says the z component of $\overline{\nabla} \times \overline{V}$ at a point is the line integral of $\overline{V}$ on a loop around that point, divided by the area of that loop, *in the limit as the loop becomes vanishingly small.* Thus, the curl does not directly tell us anything about the circulation on a macroscopic scale.

The situations in Figure 2.8 can now be understood. If the "curved" field shown in Figure 2.10(a) has a magnitude that drops off as $1/r$, exactly enough to compensate for the increase in path length as r increases, then the integral around the closed differential path shown in the figure will be zero. Thus, the curl at that point is also zero. If the magnitude of the "straight" vector field in Figure 2.10(b) varies as indicated by the line density, the integral around the differential path shown cannot be zero and the field will have a nonzero curl.

We derived Equation 2.61 in two dimensions and only picked out the z-component of the curl. The generalization of this result to three dimensions, and for any orientation of the differential loop, is straightforward and is given by

$$\lim_{C \to 0} \oint_C d\bar{r} \cdot \overline{V} = \left(\overline{\nabla} \times \overline{V} \right) \cdot \lim_{S \to 0} \int_S d\overline{\sigma}. \qquad (2.63)$$

In this equation, S is still the area enclosed by the path C. Note that the direction of $d\overline{\sigma}$ is determined by the direction of C and a right-hand convention.

2.3.4   Differential Operator Identities

Subscript/summation notation greatly facilitates the derivation of differential operator identities. The relations presented in this section are similar to the vector identities discussed in Chapter 1, except now care must be taken to obey the rules of differential calculus. As with the vector identities, a Cartesian coordinate system is used for the derivations, but the final results are expressed in coordinate-independent vector notation.

Example 2.1   Consider the operator expression, $\overline{\nabla} \cdot (\overline{\nabla}\Phi)$. To translate this into subscript/summation notation, make the substitution

$$\overline{\nabla} = \hat{\mathbf{e}}_i \frac{\partial}{\partial x_i}. \qquad (2.64)$$

The two $\overline{\nabla}$ operators in the original expression must be written using independent subscripts:

$$\overline{\nabla} \cdot \left(\overline{\nabla}\Phi\right) = \hat{\mathbf{e}}_i \frac{\partial}{\partial x_i} \cdot \left(\hat{\mathbf{e}}_j \frac{\partial}{\partial x_j} \Phi\right). \tag{2.65}$$

Because the basis vectors of a Cartesian system are position independent, $\partial \hat{\mathbf{e}}_j / \partial x_i = 0$, and Equation 2.65 becomes

$$\begin{aligned}
\overline{\nabla} \cdot \left(\overline{\nabla}\Phi\right) &= (\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j) \frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial x_j} \Phi\right) \\
&= \delta_{ij} \frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial x_j} \Phi\right) \\
&= \frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial x_i} \Phi\right) \\
&= \left(\frac{\partial^2}{\partial^2 x_1} + \frac{\partial^2}{\partial^2 x_2} + \frac{\partial^2}{\partial^2 x_3}\right)\Phi. \tag{2.66}
\end{aligned}$$

In the last line, we have written out the implied summation explicitly to show you how it is working in this case. Equation 2.66 can be returned to vector notation by defining the *Laplacian* operator ∇^2 as

$$\nabla^2 \equiv \overline{\nabla} \cdot \overline{\nabla} = \frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial x_i}\right) = \left(\frac{\partial^2}{\partial^2 x_1} + \frac{\partial^2}{\partial^2 x_2} + \frac{\partial^2}{\partial^2 x_3}\right)\Phi, \tag{2.67}$$

so that

$$\overline{\nabla} \cdot \left(\overline{\nabla}\Phi\right) = \nabla^2\Phi. \tag{2.68}$$

Example 2.2   Consider the expression $\overline{\nabla} \times \overline{\nabla} \times \overline{\mathbf{V}}$, which is the curl of the curl of $\overline{\mathbf{V}}$. This quantity arises when developing the electromagnetic wave equation from Maxwell's equations. To write this in subscript notation in a Cartesian coordinate system, we need to use two Levi-Civita symbols:

$$\overline{\nabla} \times \overline{\nabla} \times \overline{\mathbf{V}} = \frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial x_r} V_s \epsilon_{rsj}\right) \hat{\mathbf{e}}_k \epsilon_{ijk}. \tag{2.69}$$

Equation 2.69 can be manipulated as follows:

$$\begin{aligned}
\overline{\nabla} \times \overline{\nabla} \times \overline{\mathbf{V}} &= \frac{\partial}{\partial x_i} \left(\frac{\partial V_s}{\partial x_r}\right) \hat{\mathbf{e}}_k \epsilon_{rsj} \epsilon_{ijk} \\
&= -\frac{\partial}{\partial x_i} \left(\frac{\partial V_s}{\partial x_r}\right) \hat{\mathbf{e}}_k \epsilon_{rsj} \epsilon_{ikj} \\
&= -\frac{\partial}{\partial x_i} \left(\frac{\partial V_s}{\partial x_r}\right) \hat{\mathbf{e}}_k \left(\delta_{ri}\delta_{sk} - \delta_{rk}\delta_{si}\right)
\end{aligned}$$

$$= \frac{\partial}{\partial x_i}\left(\frac{\partial V_i}{\partial x_k}\right)\hat{\mathbf{e}}_k - \frac{\partial}{\partial x_i}\left(\frac{\partial V_k}{\partial x_i}\right)\hat{\mathbf{e}}_k$$

$$= \hat{\mathbf{e}}_k\frac{\partial}{\partial x_k}\left(\frac{\partial V_i}{\partial x_i}\right) - \frac{\partial}{\partial x_i}\left(\frac{\partial(V_k\hat{\mathbf{e}}_k)}{\partial x_i}\right). \qquad (2.70)$$

Finally, the right-hand side of Equation 2.70 is converted back into vector notation to obtain the result

$$\overline{\nabla} \times \overline{\nabla} \times \overline{\mathbf{V}} = \overline{\nabla}\left(\overline{\nabla} \cdot \overline{\mathbf{V}}\right) - \nabla^2\overline{\mathbf{V}}. \qquad (2.71)$$

Notice that the Laplacian operator can act on both vector and scalar fields. In Equation 2.68, the Laplacian operates on a scalar field to product a scalar. In Equation 2.71, it operates on a vector field to produce a vector.

2.4   INTEGRAL DEFINITIONS OF THE DIFFERENTIAL OPERATORS

In Equations 2.25, 2.27, and 2.29, we provided expressions for calculating the gradient, divergence, and curl. Each of these relations is valid only in a Cartesian coordinate system, and all are in terms of spatial derivatives of the field. Integral definitions of each of the differential operators also exist. We already derived one such definition for the curl in Equation 2.63. In this section, we present similar definitions for the gradient and divergence, as well as an alternate definition for the curl. Their derivations, which are all similar to the derivation of Equation 2.63, are in most introductory calculus texts. We present only the results here.

The gradient of a scalar field at a particular point can be generated from

$$\overline{\nabla}\Phi = \lim_{S,V\to 0}\frac{\oint_S \Phi\, d\overline{\sigma}}{\int_V d\tau}, \qquad (2.72)$$

where V is a volume that includes the point of interest, and S is the closed surface that surrounds V. Both S and V must be shrunk to infinitesimal size for this equation to hold.

To get the divergence of a vector field at a point, we integrate the vector field over an infinitesimal surface S that encloses that point, and divide by the infinitesimal volume:

$$\overline{\nabla} \cdot \overline{\mathbf{A}} = \lim_{S,V\to 0}\frac{\oint_S d\overline{\sigma} \cdot \overline{\mathbf{A}}}{\int_V d\tau}. \qquad (2.73)$$

We already derived the integral definition

$$\lim_{C\to 0}\oint_C d\overline{\mathbf{r}} \cdot \overline{\mathbf{V}} = \left(\overline{\nabla} \times \overline{\mathbf{V}}\right) \cdot \lim_{S\to 0}\int_S d\overline{\sigma}, \qquad (2.74)$$

which generates the curl. This definition is a bit clumsy because it requires the calculation of three different integrals, each with a different orientation of S, to get all three components of the curl. The following integral definition does not have this

problem, but it uses an uncommon form of the surface integral:

$$\overline{\nabla} \times \overline{A} = \lim_{S,V \to 0} \frac{\oint_S d\overline{\sigma} \times \overline{A}}{\int_V d\tau}.$$ (2.75)

2.5   THE THEOREMS

The differential operators give us information about the variation of vector and scalar fields on an infinitesimal scale. To apply them on a macroscopic scale, we need to introduce four important theorems. Gauss's Theorem, Green's Theorem, Stokes's Theorem, and Helmholtz's Theorem can be derived directly from the integral definitions of the differential operators. We give special attention to the proof and discussion of Helmholtz's Theorem because it is not covered adequately in many texts.

2.5.1   Gauss's Theorem

Gauss's Theorem is derived from Equation 2.73, written in a slightly different form:

$$\overline{\nabla} \cdot \overline{A}\, d\tau = \lim_{S \to 0} \oint_S d\overline{\sigma} \cdot \overline{A}.$$ (2.76)

In this equation, the closed surface S completely surrounds the volume $d\tau$, which we have written as an infinitesimal.

Equation 2.76 can be applied to two adjacent differential volumes $d\tau_1$ and $d\tau_2$ that have a common surface, as shown in Figure 2.11:

$$\overline{\nabla} \cdot \overline{A}\, d\tau_1 + \overline{\nabla} \cdot \overline{A}\, d\tau_2 = \oint_{S_1} d\overline{\sigma} \cdot \overline{A} + \oint_{S_2} d\overline{\sigma} \cdot \overline{A}.$$ (2.77)

The contributions to the surface integral from the common surfaces cancel out, as depicted in the figure, so Equation 2.77 can be written as

$$\overline{\nabla} \cdot \overline{A}\, d\tau_1 + \overline{\nabla} \cdot \overline{A}\, d\tau_2 = \oint_{S_{1+2}} d\overline{\sigma} \cdot \overline{A},$$ (2.78)

$$\overline{A} \cdot d\overline{\sigma}_1 + \overline{A} \cdot d\overline{\sigma}_2 = 0$$
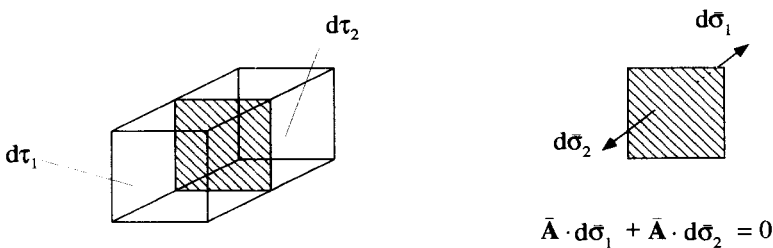
Figure 2.11   The Sum of Two Differential Volumes
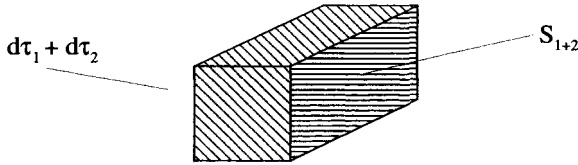
Figure 2.12   The Sum of Two Differential Volumes

where S_{1+2} is the outside surface enclosing both $d\tau_1$ and $d\tau_2$, as shown in Figure 2.12. We can continue this process of adding up contiguous differential volume to form an arbitrary finite volume V enclosed by a closed surface S. The result,

$$\int_V d\tau\, \overline{\nabla} \cdot \overline{\mathbf{A}} = \oint_S d\overline{\sigma} \cdot \overline{\mathbf{A}}, \qquad (2.79)$$

is called *Gauss's Theorem*.

2.5.2   Green's Theorem

Green's Theorem takes two forms and follows directly from Gauss's Theorem and some vector operator manipulations. Start by considering the expression $\overline{\nabla} \cdot (u\overline{\nabla} v)$, where u and v are both scalar fields. Using a vector operator identity, which you will prove in one of this chapter's exercises, this expression can be rewritten as

$$\overline{\nabla} \cdot \left(u\overline{\nabla} v\right) = \overline{\nabla} u \cdot \overline{\nabla} v + u\nabla^2 v. \qquad (2.80)$$

Switching u and v gives

$$\overline{\nabla} \cdot \left(v\overline{\nabla} u\right) = \overline{\nabla} v \cdot \overline{\nabla} u + v\nabla^2 u. \qquad (2.81)$$

When Equation 2.81 is subtracted from Equation 2.80, the result is

$$\overline{\nabla} \cdot \left(u\overline{\nabla} v\right) - \overline{\nabla} \cdot \left(v\overline{\nabla} u\right) = u\nabla^2 v - v\nabla^2 u. \qquad (2.82)$$

Finally, we integrate both sides of Equation 2.82 over a volume V and apply Gauss's Theorem, to generate one form of *Green's Theorem*:

$$\oint_S d\overline{\sigma} \cdot \left(u\overline{\nabla} v - v\overline{\nabla} u\right) = \int_V d\tau \left[u\nabla^2 v - v\nabla^2 u\right]. \qquad (2.83)$$

In this expression, the closed surface S surrounds the volume V. The same process applied directly to Equation 2.80 produces a second form of Green's Theorem:

$$\oint_S d\overline{\sigma} \cdot \left(u\overline{\nabla} v\right) = \int_V d\tau \left[\overline{\nabla} u \cdot \overline{\nabla} v + u\nabla^2 v\right]. \qquad (2.84)$$

2.5.3 Stokes's Theorem

Stokes's Theorem derives from Equation 2.74,

$$\left(\overline{\nabla} \times \overline{A}\right) \cdot d\overline{\sigma} = \lim_{C \to 0} \oint_C d\overline{r} \cdot \overline{A}, \tag{2.85}$$

where the path C encloses the differential surface $d\overline{\sigma}$ in the right-hand sense.

The development of Stokes's theorem follows steps similar to those for Gauss's theorem. Equation 2.85 is applied to two adjacent differential surfaces that have a common border, as shown in Figure 2.13. The result is

$$\left(\overline{\nabla} \times \overline{A}\right) \cdot d\overline{\sigma}_1 + \left(\overline{\nabla} \times \overline{A}\right) \cdot d\overline{\sigma}_2 = \oint_{C_{1+2}} d\overline{r} \cdot \overline{A} \tag{2.86}$$

where the path C_{1+2} is the closed path surrounding both $d\sigma_1$ and $d\sigma_2$. The line integrals along the common edge of C_1 and C_2 cancel exactly. Any number of these differential areas can be added up to give an arbitrary surface S and the closed contour C which surrounds S. The result is *Stokes's Theorem*:

$$\int_S d\overline{\sigma} \cdot \left(\overline{\nabla} \times \overline{A}\right) = \oint_C d\overline{r} \cdot \overline{A}. \tag{2.87}$$

There is an important consequence of Stokes's Theorem for vector fields that have zero curl. Such a field can always be derived from a scalar potential. That is to say, if $\overline{\nabla} \times \overline{A} = 0$ everywhere, then there exists a scalar function $\Phi(\overline{r})$ such that $\overline{A} = -\overline{\nabla}\Phi$. To see this, consider the two points 1 and 2 and two arbitrary paths between them, Path A and Path B, as shown in Figure 2.14. A closed line integral can be formed by combining Path A and the *reversal* of path B. If $\overline{\nabla} \times \overline{A} = 0$ everywhere, then Equation 2.87 lets us write

$$\int_A d\overline{r} \cdot \overline{A} - \int_B d\overline{r} \cdot \overline{A} = \oint d\overline{r} \cdot \overline{A} = 0, \tag{2.88}$$
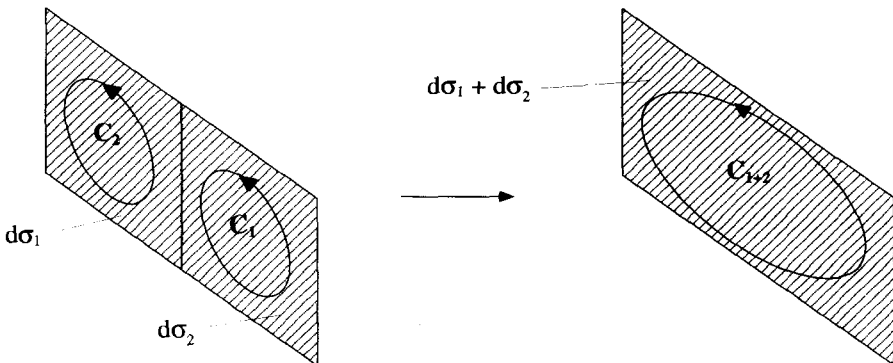
Figure 2.13   The Sum of Two Differential Surfaces
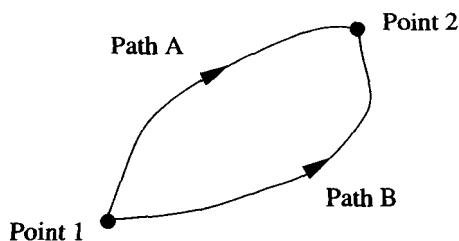
Figure 2.14   Stokes's Theorem Implies a Scalar Potential

or

$$\int_A d\bar{\mathbf{r}} \cdot \overline{\mathbf{A}} = \int_B d\bar{\mathbf{r}} \cdot \overline{\mathbf{A}}. \qquad (2.89)$$

Equation 2.89 says the line integral of $\overline{\mathbf{A}}$ between the two points is independent of the path taken. This means that it is possible to define a scalar function of position $\Phi(\bar{\mathbf{r}})$ such that its total differential is given by

$$d\Phi = -d\bar{\mathbf{r}} \cdot \overline{\mathbf{A}}. \qquad (2.90)$$

It is conventional to add the negative sign here so Φ increases as you move against the field lines of $\overline{\mathbf{A}}$. Inserting Equation 2.90 into the line integrals of Equation 2.89 shows that these integrals are both equal to

$$\int_1^2 -d\Phi = \Phi(1) - \Phi(2). \qquad (2.91)$$

Referring back to Equation 2.33, it is clear that the condition of Equation 2.90 can be rewritten as

$$\overline{\mathbf{A}} = -\overline{\nabla}\Phi. \qquad (2.92)$$

In summary, if the curl of a vector field is everywhere zero, the field is derivable from a scalar potential. The line integral of this type of vector field is always independent of the path taken. Fields of this type are often called *conservative*.

2.5.4   Helmholtz's Theorem

Helmholtz's Theorem states:

> A vector field, if it exists, is uniquely determined by specifying its divergence and curl everywhere within a region and its normal component on the closed surface surrounding that region.

There are two important parts of this statement. On one hand, it says if we have a field $\overline{V}$ that we are trying to determine, and we know the values of $\overline{\nabla} \cdot \overline{V}$ and $\overline{\nabla} \times \overline{V}$ at every point in some volume, plus the normal component of $\overline{V}$ on the surface of that volume, there is only one $\overline{V}$ that will make all this work. On the other hand, we have the disclaimer, "if it exists." This qualification is necessary because it is entirely possible to specify values for the divergence, curl, and normal component of a vector field that cannot be satisfied by any field.

To prove Helmholtz's Theorem we will assume the two vector fields $\overline{V}_1$ and $\overline{V}_2$ have the same values for the divergence, curl, and normal component. Then we show that if this is the case, the two solutions must be equal. Let $\overline{W} = \overline{V}_1 - \overline{V}_2$. Since the divergence, curl, and the dot product are all linear operators, $\overline{W}$ must have the following properties:

$$\overline{\nabla} \cdot \overline{W} = 0 \qquad \text{in the region}$$

$$\overline{\nabla} \times \overline{W} = 0 \qquad \text{in the region} \qquad\qquad (2.93)$$

$$\hat{n} \cdot \overline{W} = 0 \qquad \text{on the surface.}$$

Because $\overline{\nabla} \times \overline{W} = 0$, $\overline{W}$ can be derived from a scalar potential

$$\overline{W} = -\overline{\nabla}\Phi. \qquad\qquad (2.94)$$

Green's Theorem, in the form of Equation 2.84, is now applied with $u = v = \Phi$ to give

$$\oint_S d\overline{\sigma} \cdot \Phi\left(\overline{\nabla}\Phi\right) = \int_V d\tau \left[\Phi\overline{\nabla} \cdot \left(\overline{\nabla}\Phi\right) + \overline{\nabla}\Phi \cdot \overline{\nabla}\Phi\right]. \qquad (2.95)$$

Inserting Equation 2.94 into Equation 2.95 gives

$$\oint_S d\overline{\sigma} \cdot \Phi\overline{W} = \int_V d\tau\left(\Phi\overline{\nabla} \cdot \overline{W} - \overline{W} \cdot \overline{W}\right). \qquad (2.96)$$

Using Equations 2.93, the surface integral on the LHS and the volume integral of $\Phi\overline{\nabla} \cdot \overline{W}$ on the RHS are zero, with the result that

$$\int_V d\tau\overline{W} \cdot \overline{W} = \int_V d\tau|\overline{W}|^2 = 0. \qquad (2.97)$$

Because $|\overline{W}|^2$ is always a positive quantity, the only way to satisfy Equation 2.97 is to have $\overline{W} = 0$ everywhere. Consequently, $\overline{V}_1 = \overline{V}_2$ and *Helmholtz's Theorem* is proven.

Helmholtz's Theorem is useful for separating vectors fields into two parts, one with zero curl, and one with zero divergence. This discussion relies on two identities

$$\overline{\nabla} \cdot \left(\overline{\nabla} \times \overline{A}\right) = 0 \qquad\qquad (2.98)$$

$$\overline{\nabla} \times \overline{\nabla}\Phi = 0, \qquad\qquad (2.99)$$

which you will prove at the end of the chapter. Write $\overline{\mathbf{V}}$ as

$$\overline{\mathbf{V}} = \overline{\mathbf{V}} \times \overline{\mathbf{A}} - \overline{\mathbf{V}}\Phi. \tag{2.100}$$

Then we can write:

$$\overline{\mathbf{V}} \cdot \overline{\mathbf{V}} = -\nabla^2 \Phi$$

$$\overline{\mathbf{V}} \times \overline{\mathbf{V}} = \overline{\mathbf{V}} \times \overline{\mathbf{V}} \times \overline{\mathbf{A}}.$$

$$\hat{\mathbf{n}} \cdot \overline{\mathbf{V}} = \hat{\mathbf{n}} \cdot \left(\overline{\mathbf{V}} \times \overline{\mathbf{A}} - \overline{\mathbf{V}}\Phi\right), \tag{2.101}$$

Since the divergence, curl, and normal component are all specified if $\overline{\mathbf{A}}$ and Φ are specified, Helmholtz's Theorem says $\overline{\mathbf{V}}$ is also uniquely specified. Notice the contribution to $\overline{\mathbf{V}}$ that comes from $\overline{\mathbf{A}}$ has no divergence since $\overline{\mathbf{V}} \cdot (\overline{\mathbf{V}} \times \overline{\mathbf{A}}) = 0$. This is called the rotational or solenoidal part of the field and $\overline{\mathbf{A}}$ is called the vector potential. The portion of $\overline{\mathbf{V}}$ which arises from Φ has no curl, because $\overline{\mathbf{V}} \times \overline{\mathbf{V}}\Phi = 0$. This is called the irrotational part of the field and Φ is called the scalar potential.

EXERCISES FOR CHAPTER 2

1. Consider the two-dimensional scalar potential function

$$\Phi(x, y) = x^3 - 3y^2 x.$$

   (a) Make a plot of the equipotential contours for three positive and three negative values for Φ.
   (b) Find $\overline{\mathbf{V}}\Phi$.
   (c) Show that the $\overline{\mathbf{V}}\Phi$ field lines are given by setting $3x^2 y - y^3$ equal to a series of constants.
   (d) Plot six representative $\overline{\mathbf{V}}\Phi$ field lines. Be sure to indicate the direction of the field and comment on its magnitude.
   (e) Find $\overline{\mathbf{V}} \cdot \overline{\mathbf{V}}\Phi$, the divergence of the $\overline{\mathbf{V}}\Phi$ vector field, and show that your field lines of part (d) agree with this divergence.

2. An electric dipole $\overline{\mathbf{p}}$ is located at the origin of a Cartesian system. This dipole creates an electric potential field $\Phi(\overline{\mathbf{r}})$ given by

$$\Phi(\overline{\mathbf{r}}) = \frac{\overline{\mathbf{p}} \cdot \overline{\mathbf{r}}}{4\pi\epsilon_o |\overline{\mathbf{r}}|^3},$$
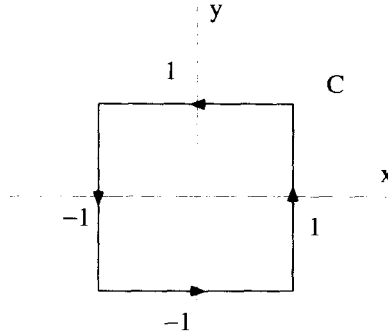
   where $\overline{\mathbf{r}}$ is the position vector, $\overline{\mathbf{r}} = x_i \hat{\mathbf{e}}_i$. Let $\overline{\mathbf{p}} = p\hat{\mathbf{e}}_3$.
   (a) Sketch the equipotential lines, $\Phi = $ constant.
   (b) Find the electric field, $\overline{\mathbf{E}} = -\overline{\mathbf{V}}\Phi$.
   (c) Sketch the electric field lines.

3. Perform the line integral

$$\oint_C d\bar{\mathbf{r}} \times \bar{\mathbf{V}},$$

where C is the contour shown below

and

(a) $\bar{\mathbf{V}} = V_o \hat{\mathbf{e}}_x$.

(b) $\bar{\mathbf{V}} = V_o \hat{\mathbf{e}}_z$.

(c) $\bar{\mathbf{V}} = \dfrac{V_o x}{\sqrt{x^2 + y^2}} \hat{\mathbf{e}}_x + \dfrac{V_o y}{\sqrt{x^2 + y^2}} \hat{\mathbf{e}}_y$,

4. Use subscript/summation notation to verify the following identities:

(a) $\bar{\mathbf{\nabla}} \cdot \left(f\bar{\mathbf{A}} \right) = f \left(\bar{\mathbf{\nabla}} \cdot \bar{\mathbf{A}} \right) + \bar{\mathbf{A}} \cdot \bar{\mathbf{\nabla}} f$.

(b) $\bar{\mathbf{\nabla}} \times \bar{\mathbf{\nabla}} \times \bar{\mathbf{A}} = \bar{\mathbf{\nabla}} \left(\bar{\mathbf{\nabla}} \cdot \bar{\mathbf{A}} \right) - \nabla^2 \bar{\mathbf{A}}$.

(c) $\bar{\mathbf{\nabla}} \times \left(f\bar{\mathbf{A}} \right) = f \left(\bar{\mathbf{\nabla}} \times \bar{\mathbf{A}} \right) + \bar{\mathbf{\nabla}} f \times \bar{\mathbf{A}}$.

(d) $\bar{\mathbf{\nabla}} \cdot \left(\bar{\mathbf{A}} \times \bar{\mathbf{B}} \right) = \bar{\mathbf{B}} \cdot \left(\bar{\mathbf{\nabla}} \times \bar{\mathbf{A}} \right) - \bar{\mathbf{A}} \cdot \left(\bar{\mathbf{\nabla}} \times \bar{\mathbf{B}} \right)$.

(e) $\bar{\mathbf{A}} \times \left(\bar{\mathbf{\nabla}} \times \bar{\mathbf{B}} \right) = \left(\bar{\mathbf{\nabla}} \bar{\mathbf{B}} \right) \cdot \bar{\mathbf{A}} - \left(\bar{\mathbf{A}} \cdot \bar{\mathbf{\nabla}} \right) \bar{\mathbf{B}}$.

(f) $\bar{\mathbf{\nabla}}(fg) = f\bar{\mathbf{\nabla}} g + g\bar{\mathbf{\nabla}} f$.

(g) $\bar{\mathbf{\nabla}} \left(\bar{\mathbf{A}} \cdot \bar{\mathbf{B}} \right) = \bar{\mathbf{A}} \times \left(\bar{\mathbf{\nabla}} \times \bar{\mathbf{B}} \right) + \bar{\mathbf{B}} \times \left(\bar{\mathbf{\nabla}} \times \bar{\mathbf{A}} \right) + \left(\bar{\mathbf{A}} \cdot \bar{\mathbf{\nabla}} \right) \bar{\mathbf{B}} + \left(\bar{\mathbf{B}} \cdot \bar{\mathbf{\nabla}} \right) \bar{\mathbf{A}}$.

(h) $\bar{\mathbf{\nabla}} \times \bar{\mathbf{\nabla}} f = 0$.

(i) $\bar{\mathbf{\nabla}} \cdot \left(\bar{\mathbf{\nabla}} \times \bar{\mathbf{A}} \right) = 0$.

5. Calculate the work done by following a straight line path from the Cartesian point $(1, 1)$ to $(3, 3)$, if the force exerted is given by

$$\bar{\mathbf{F}} = (x - y)\hat{\mathbf{e}}_x + (x + y)\hat{\mathbf{e}}_y.$$

Can this force be derived from a scalar potential? Pick any other path that goes from $(1, 1)$ to $(3, 3)$ and calculate the work done.

6. Let a vector field $\overline{\mathbf{V}}$ be given by

$$\overline{\mathbf{V}} = x\hat{\mathbf{e}}_x + y\hat{\mathbf{e}}_y.$$

   (a) Use Stokes's Theorem to evaluate the line integral,

$$\oint_C d\bar{\mathbf{r}} \cdot \overline{\mathbf{V}},$$

   where C is the unit circle in the xy-plane centered at $(0,0)$.
   (b) Use Stokes's Theorem to determine the same line integral if C is a unit circle centered at $(1,0)$.
   (c) Evaluate the two integrals above by actually performing the line integrals themselves.

7. Show that if $\overline{\mathbf{B}} = \overline{\nabla} \times \overline{\mathbf{A}}$, then for any closed surface S,

$$\oint_S d\overline{\boldsymbol{\sigma}} \cdot \overline{\mathbf{B}} = 0.$$

8. Use the differential operator theorems to evaluate

$$\oint_S d\overline{\boldsymbol{\sigma}} \cdot \overline{\mathbf{V}},$$

   where S is the surface of a sphere of radius R centered at the origin, and $\overline{\mathbf{V}}$ is
   (a) $\overline{\mathbf{V}} = x\hat{\mathbf{e}}_x + y\hat{\mathbf{e}}_y + z\hat{\mathbf{e}}_z$.
   (b) $\overline{\mathbf{V}} = x^3\hat{\mathbf{e}}_x + y^3\hat{\mathbf{e}}_y + z^3\hat{\mathbf{e}}_z$.
   (c) $\overline{\mathbf{V}} = x^5\hat{\mathbf{e}}_x + y^5\hat{\mathbf{e}}_y + z^5\hat{\mathbf{e}}_z$.

9. Use an appropriate vector theorem to evaluate

$$\oint_C d\bar{\mathbf{r}} \cdot \overline{\mathbf{V}},$$

   where C is any closed path and

$$\overline{\mathbf{V}} = (2y^2 - 3x^2 y)\hat{\mathbf{e}}_x + (4xy - x^3)\hat{\mathbf{e}}_y.$$

10. A force field is given by the relationship

$$\overline{\mathbf{F}} = (2x + 2y)\hat{\mathbf{e}}_x + (2y + 2x)\hat{\mathbf{e}}_y.$$

   (a) What is $\overline{\nabla} \cdot \overline{\mathbf{F}}$?
   (b) What is $\overline{\nabla} \times \overline{\mathbf{F}}$?
   (c) Sketch the force field lines.

(d) This is a conservative field. What is its potential?

(e) Sketch the equipotentials.

11. The continuity equation for a fluid is

$$\overline{\nabla} \cdot \overline{\mathbf{J}} + \frac{\partial \rho}{\partial t} = 0,$$

where $\overline{\mathbf{J}} = \rho \overline{\mathbf{V}}$.

(a) Show $\overline{\nabla} \cdot \overline{\mathbf{V}} = 0$ if the fluid is incompressible (constant density).

(b) Apply Gauss's Theorem to the continuity equation and interpret the result.

12. Prove these integral forms of the differential operators:

(a) $\overline{\nabla} \Phi = \lim\limits_{S,V \to 0} \dfrac{\oint_S \Phi \, d\overline{\sigma}}{\int_V d\tau}$.

(b) $\overline{\nabla} \cdot \overline{\mathbf{A}} = \lim\limits_{S,V \to 0} \dfrac{\oint_S d\overline{\sigma} \cdot \overline{\mathbf{A}}}{\int_V d\tau}$.

(c) $\overline{\nabla} \times \overline{\mathbf{A}} = \lim\limits_{S,V \to 0} \dfrac{\oint_S d\overline{\sigma} \times \overline{\mathbf{A}}}{\int_V d\tau}$.

13. Maxwell's equations in vacuum are

$$\overline{\nabla} \times \overline{\mathbf{E}} = -\frac{\partial \overline{\mathbf{B}}}{\partial t}$$

$$\overline{\nabla} \times \overline{\mathbf{B}} = \mu_o \epsilon_o \frac{\partial \overline{\mathbf{E}}}{\partial t}$$

$$\overline{\nabla} \cdot \overline{\mathbf{E}} = 0$$

$$\overline{\nabla} \cdot \overline{\mathbf{B}} = 0.$$

Manipulate these using subscript/summation notation to obtain the wave equation in vacuum:

$$\nabla^2 \overline{\mathbf{E}} = \mu_o \epsilon_o \frac{\partial^2 \overline{\mathbf{E}}}{\partial t^2}.$$

3

CURVILINEAR COORDINATE SYSTEMS

Up to this point, our discussions of vector, differential, and integral operations have been limited to Cartesian coordinate systems. While conceptually simple, these systems often fail to utilize the natural symmetry of certain problems. Consider the electric field vector created by a point charge q, located at the origin of a Cartesian system. Using Cartesian basis vectors, this field is

$$\overline{E} = \frac{q}{4\pi\epsilon_0} \frac{x\hat{e}_x + y\hat{e}_y + z\hat{e}_z}{(x^2 + y^2 + z^2)^{3/2}}. \tag{3.1}$$

In contrast, a spherical system, described by the coordinates (r, θ, ϕ), fully exploits the symmetry of this field and simplifies Equation 3.1 to

$$\overline{E} = \frac{q}{4\pi\epsilon_0} \frac{\hat{e}_r}{r^2}. \tag{3.2}$$

The spherical system belongs to the class of *curvilinear* coordinate systems. Basis vectors of a curvilinear system are orthonormal, just like those of Cartesian systems, but their directions can be functions of position.

This chapter generalizes the concepts of the previous chapters to include curvilinear coordinate systems. The two most common systems, spherical and cylindrical, are described first, in order to provide a framework for the more abstract discussion of *generalized* curvilinear coordinates that follows.

3.1 THE POSITION VECTOR

The position vector $\overline{r}(P)$ associated with a point P describes the offset of P from the origin of the coordinate system. It has a magnitude equal to the distance from the origin to P, and a direction that points from the origin to P.
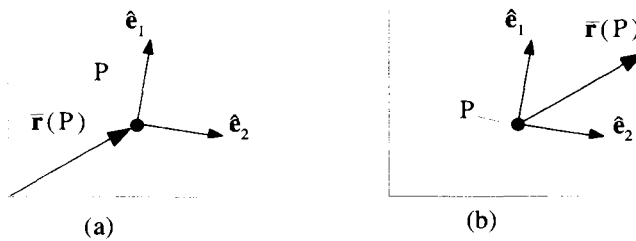
44

Figure 3.1   The Position Vector

It seems natural to draw the position vector between the origin and P, as shown in Figure 3.1(a). While this is fine for Cartesian coordinate systems, it can lead to difficulties in curvilinear systems. The problems arise because of the position dependence of the curvilinear basis vectors. When we draw any vector, we must always be careful to specify where it is located. If we did not do this, it would not be clear how to expand the vector in terms of the basis vectors. To get around this difficulty, both the vector and the basis vector should be drawn emanating from the point in question. The curvilinear vector components are then easily obtained by projecting the vector onto the basis vectors at that point. Consequently, to determine the components of the position vector, it is better to draw it, as well as the basis vectors, emanating from P. This is shown in Figure 3.1(b). There are situations, however, when it is better to draw the position vector emanating from the origin. For example, line integrals, such as the one shown in Figure 2.2, are best described in this way, because then the tip of the position vector follows the path of the integration. We will place the position vector as shown in Figure 3.1(a) or (b), depending upon which is most appropriate for the given situation.

In Cartesian coordinates, the expression for the position vector is intuitive and simple:

$$\bar{\mathbf{r}} = r_i \hat{\mathbf{e}}_i = x_i \hat{\mathbf{e}}_i. \tag{3.3}$$

The components (r_1, r_2, r_3) are easily identified as the Cartesian coordinates (x_1, x_2, x_3). Formally, r_1 is obtained by dot multiplying the position vector $\bar{\mathbf{r}}$ by the basis vector $\hat{\mathbf{e}}_1$:

$$r_1 \equiv \hat{\mathbf{e}}_1 \cdot \bar{\mathbf{r}} = x_1. \tag{3.4}$$

While this may seem overly pedantic here, this technique can be used to find the vector components for any vector in any orthogonal coordinate system.

3.2   THE CYLINDRICAL SYSTEM

The coordinates of a point P described in a cylindrical system are (ρ, ϕ, z). The equations

$$x = \rho \cos \phi$$
$$y = \rho \sin \phi \qquad (3.5)$$
$$z = z$$

and the corresponding inverse equations

$$\rho = \sqrt{x^2 + y^2}$$
$$\phi = \tan^{-1}\left(y/x\right) \qquad (3.6)$$
$$z = z$$

govern the relationship between cylindrical coordinates and the coordinates of a superimposed Cartesian system, as shown in Figure 3.2(a).

The unit basis vectors for the cylindrical system are shown in Figure 3.2(b). Each basis vector points in the direction that P moves when the corresponding coordinate is increased. For example, the direction of $\hat{\mathbf{e}}_\rho$ is found by watching how P moves as ρ is increased. This method can be used to determine the directions of basis vectors for any set of coordinates. Unlike the Cartesian system, the cylindrical basis vectors are not fixed. As the point P moves, the directions of $\hat{\mathbf{e}}_\rho$ and $\hat{\mathbf{e}}_\phi$ both change. Also notice that if P lies exactly on the z-axis, i.e., $\rho = 0$, the directions of $\hat{\mathbf{e}}_\rho$ and $\hat{\mathbf{e}}_\phi$ are undefined.

The cylindrical coordinates, taken in the order (ρ, ϕ, z), form a right-handed system. If you align your right hand along $\hat{\mathbf{e}}_\rho$, and then curl your fingers to point in the direction of $\hat{\mathbf{e}}_\phi$, your thumb will point in the $\hat{\mathbf{e}}_z$ direction. The basis vectors are also orthonormal since

$$\hat{\mathbf{e}}_\rho \cdot \hat{\mathbf{e}}_\phi = \hat{\mathbf{e}}_\rho \cdot \hat{\mathbf{e}}_z = \hat{\mathbf{e}}_z \cdot \hat{\mathbf{e}}_\phi = 0$$
$$\hat{\mathbf{e}}_\rho \cdot \hat{\mathbf{e}}_\rho = \hat{\mathbf{e}}_\phi \cdot \hat{\mathbf{e}}_\phi = \hat{\mathbf{e}}_z \cdot \hat{\mathbf{e}}_z = 1. \qquad (3.7)$$

The position vector expressed in cylindrical coordinates is

$$\bar{\mathbf{r}} = (\bar{\mathbf{r}} \cdot \hat{\mathbf{e}}_\rho)\hat{\mathbf{e}}_\rho + (\bar{\mathbf{r}} \cdot \hat{\mathbf{e}}_\phi)\hat{\mathbf{e}}_\phi + (\bar{\mathbf{r}} \cdot \hat{\mathbf{e}}_z)\hat{\mathbf{e}}_z. \qquad (3.8)$$

(a)   (b)
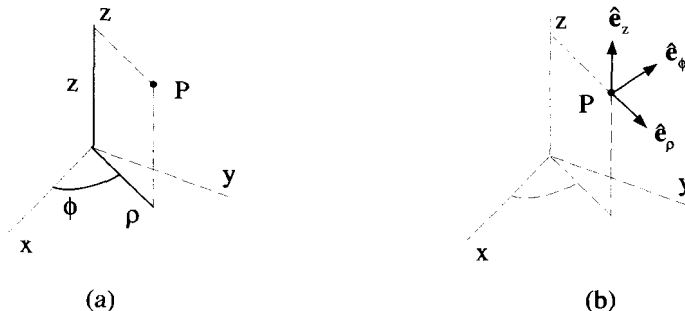
Figure 3.2   The Cylindrical System
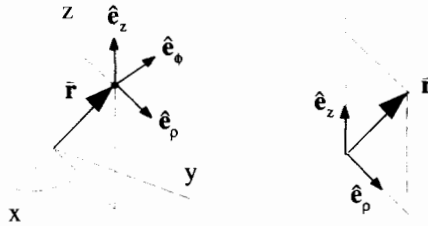
Figure 3.3 The Position Vector in a Cylindrical System

Notice that $\hat{e}_\phi$ is always perpendicular to $\bar{r}$, as shown in Figure 3.3, so Equation 3.8 reduces to

$$\bar{r} = r_\rho \hat{e}_\rho + r_z \hat{e}_z. \qquad (3.9)$$

The two-dimensional version of the cylindrical system, with only the (ρ, ϕ) coordinates, is called a polar system. This system, shown in Figure 3.4(a), has basis vectors $\hat{e}_\rho$ and $\hat{e}_\phi$. The position vector, shown in Figure 3.4(b), has only a ρ-component and is expressed as

$$\bar{r} = \rho \hat{e}_\rho. \qquad (3.10)$$

Remember an arbitrary vector $\overline{V}$, unlike the position vector, can have both ρ- and ϕ-components, as shown in Figure 3.5.

(a)   (b)
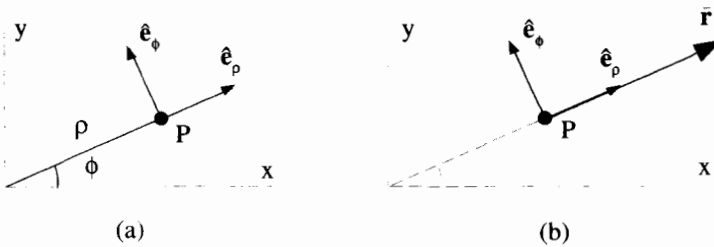
Figure 3.4 The Polar System
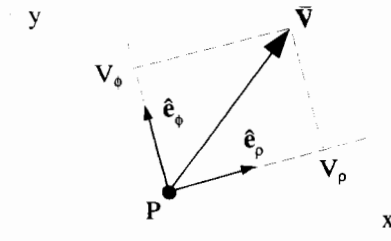
Figure 3.5 Polar Components of a Vector

3.3   THE SPHERICAL SYSTEM

The three coordinates (r, θ, ϕ) describe a point in a spherical coordinate system. Their relationship to a set of Cartesian coordinates is shown in Figure 3.6(a). The equations

$$x = r \sin \theta \cos \phi$$
$$y = r \sin \theta \sin \phi \qquad (3.11)$$
$$z = r \cos \theta$$

and the inverses

$$r = \sqrt{x^2 + y^2 + z^2}$$
$$\phi = \cos^{-1} \left(\frac{x}{\sqrt{x^2 + y^2}} \right) \qquad (3.12)$$
$$\theta = \cos^{-1} \left(\frac{z}{\sqrt{x^2 + y^2 + z^2}} \right)$$

allow conversion between the coordinates of the two systems.

The unit basis vectors for the spherical system are shown in Figure 3.6(b). As with cylindrical coordinates, we obtain the direction of each basis vector by increasing the associated coordinate and watching how P moves. Notice how the basis vectors change with the position of the point P. If P lies on the z-axis the directions for $\hat{\mathbf{e}}_\theta$ and $\hat{\mathbf{e}}_\phi$ are not defined. If P lies at the origin $\hat{\mathbf{e}}_r$ is also not defined.

The spherical system, with the coordinates in the order (r, θ, ϕ), is right-handed, just as in both the Cartesian and cylindrical systems. It is also an orthonormal system because

$$\hat{\mathbf{e}}_r \cdot \hat{\mathbf{e}}_\theta = \hat{\mathbf{e}}_r \cdot \hat{\mathbf{e}}_\phi = \hat{\mathbf{e}}_\theta \cdot \hat{\mathbf{e}}_\phi = 0$$
$$\hat{\mathbf{e}}_r \cdot \hat{\mathbf{e}}_r = \hat{\mathbf{e}}_\theta \cdot \hat{\mathbf{e}}_\theta = \hat{\mathbf{e}}_\phi \cdot \hat{\mathbf{e}}_\phi = 1. \qquad (3.13)$$

(a)   (b)
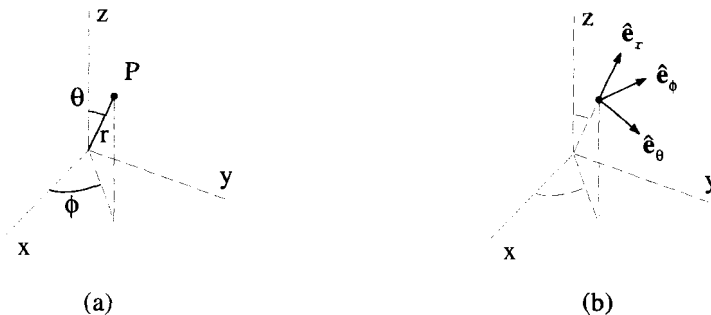
Figure 3.6   The Spherical System
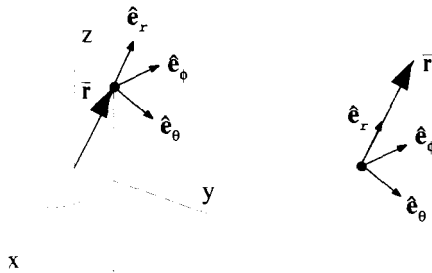
Figure 3.7   The Position Vector in Spherical Coordinates

The position vector, shown in Figure 3.7, is expressed in the spherical system as

$$\bar{r} = (\bar{r} \cdot \hat{e}_\rho)\hat{e}_\rho + (\bar{r} \cdot \hat{e}_\theta)\hat{e}_\theta + (\bar{r} \cdot \hat{e}_\phi)\hat{e}_\phi. \tag{3.14}$$

Because $\bar{r}$ is always perpendicular to $\hat{e}_\theta$ and $\hat{e}_\phi$, Equation 3.14 simplifies to

$$\bar{r} = r\hat{e}_r. \tag{3.15}$$

3.4   GENERAL CURVILINEAR SYSTEMS

Although the most common, cylindrical and spherical coordinate systems are just two examples of the larger family of curvilinear systems. A system is classified as curvilinear if it has orthonormal, but not necessarily constant, basis vectors. Other more esoteric curvilinear systems include the toroidal, hyperbolic, and elliptical systems. Instead of individually working out the vector operations of the previous chapter for each of these systems, we present a general approach that can tackle any curvilinear geometry.

3.4.1   Coordinates, Basis Vectors, and Scale Factors

The coordinates (q_1, q_2, q_3) and corresponding basis vectors $(\hat{q}_1, \hat{q}_2, \hat{q}_3)$ will be used to represent any generic curvilinear system, as shown in Figure 3.8. Because these

Figure 3.8   Curvilinear Coordinates and Basis Vectors

basis vectors are functions of position, we should always be careful to draw them emanating from a particular point, as we mentioned earlier in this chapter.

In both the cylindrical and spherical coordinate systems, a set of equations existed which related these coordinates to a "standard" set of Cartesian coordinates. For the general case, we write these equations as
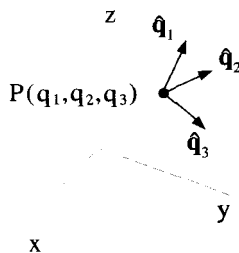
$$x_i = x_i(q_1, q_2, q_3) \tag{3.16}$$

$$q_i = q_i(x_1, x_2, x_3), \tag{3.17}$$

where the subscript notation has crept in to keep things concise. In both these equations, the subscript i takes on the values $(1, 2, 3)$. The variables x_i always represent Cartesian coordinates, while the q_i are general curvilinear coordinates.

An expression for $\hat{\mathbf{q}}_i$, the unit basis vector associated with the coordinate q_i, can be constructed by increasing q_i, watching how the position vector changes, and then normalizing:

$$\hat{\mathbf{q}}_i = \frac{\partial \bar{\mathbf{r}}/\partial q_i}{h_i}, \tag{3.18}$$

where $h_i = |\partial \bar{\mathbf{r}}/\partial q_i|$. This equation is a little confusing, because there actually is no sum over the i index on the RHS, even though it appears twice. This is subtly implied by the notation, because there is an i subscript on the LHS. The h_i, which are sometimes called scale factors, force the basis vectors to have unit length. They can be written in terms of the curvilinear coordinates. To see this, write the position vector in terms of its Cartesian components, which in turn are written as functions of the curvilinear coordinates:

$$\bar{\mathbf{r}} = x_j(q_1, q_2, q_3)\hat{\mathbf{e}}_j. \tag{3.19}$$

Therefore,

$$\frac{\partial \bar{\mathbf{r}}}{\partial q_i} = \frac{\partial x_j(q_1, q_2, q_3)}{\partial q_i}\hat{\mathbf{e}}_j \tag{3.20}$$

and

$$h_i = \left|\frac{\partial \bar{\mathbf{r}}}{\partial q_i}\right| = \sqrt{\left(\frac{\partial x_1}{\partial q_i}\right)^2 + \left(\frac{\partial x_2}{\partial q_i}\right)^2 + \left(\frac{\partial x_3}{\partial q_i}\right)^2}. \tag{3.21}$$

The physical interpretation of the scale factors is quite simple. For a change dq_1 of the coordinate q_1, the position vector changes by a distance of $|dq_1 h_1|$. Therefore, using Equation 3.18, the displacement vector can be written in the curvilinear system as

$$d\bar{\mathbf{r}} = \frac{\partial \bar{\mathbf{r}}}{\partial q_i}dq_i$$

$$= dq_i h_i(q_1, q_2, q_3)\hat{\mathbf{q}}_i, \tag{3.22}$$

where there now is a sum over the index i on the RHS because there is no subscript on the LHS. Since the h_i factors can change with position, a differential volume element in a curvilinear system is not necessarily a square cube as it is in Cartesian systems. Instead, as discussed in the next section, the sides of the differential volume in a curvilinear system vary in length and can pick up curvature.

3.4.2 Differential Geometry

Figure 3.9 depicts a differential surface enclosing an infinitesimal volume in a curvilinear system. This figure will be the basis for the derivation, in general curvilinear coordinates, of the divergence and curl, as well as surface and volume integrals.

The volume is formed by choosing a starting point (q_1, q_2, q_3) and then constructing seven other vertices by moving away from this point with small changes of coordinates dq_1, dq_2, or dq_3. In the differential limit, the lengths along each edge of this "warped cube" are given by the appropriate dq_i times its scale factor. The scale factor is evaluated at a set of coordinates that corresponds to their initial value on the edge. If the coordinate q_i stays at q_i all along an edge, it is set equal to q_i. If the coordinate q_i is equal to $q_i + dq_i$ all along an edge, it is set equal to $q_i + dq_i$. If the coordinate q_i
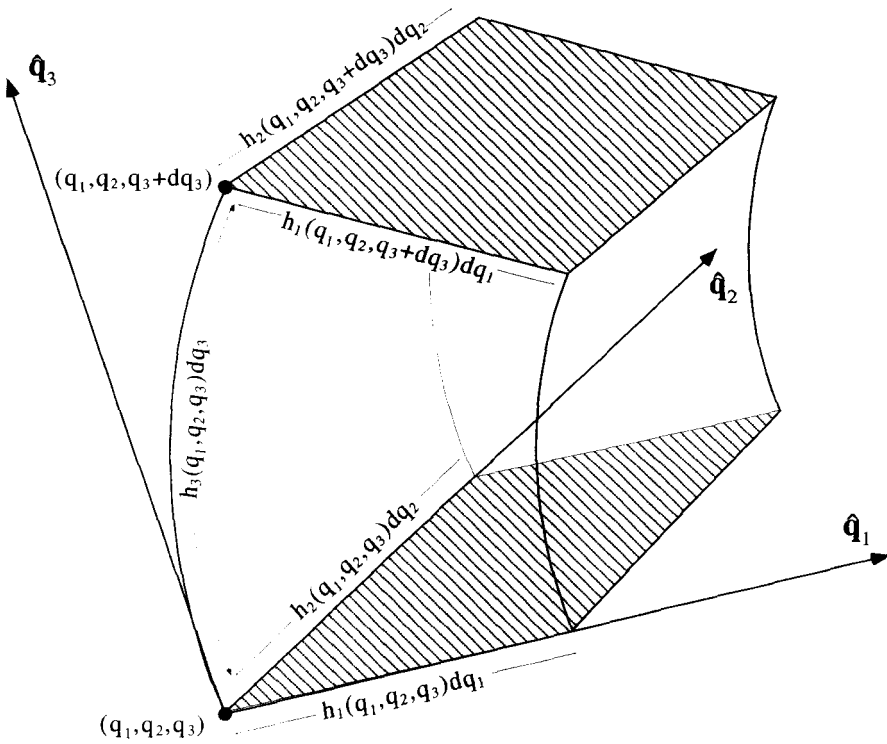

Figure 3.9   Differential Volume of a Curvilinear Coordinate System

goes from q_i to $q_i + dq_i$ on an edge, it is set equal to q_i. This is a somewhat cavalier way to treat the position dependence of these factors, although it does give all the correct results for our derivations. A more rigorous approach, which evaluates the mean value of the scale factors on each edge, is presented in Appendix B in a detailed derivation of the curvilinear curl.

Following this approach, the differential element's volume is simply

$$d
\tau = dq_1 dq_2 dq_3 h_1 h_2 h_3 \Big|_{(q_1, q_2, q_3)}, \tag{3.23}$$

where the h_i's are all evaluated at the point (q_1, q_2, q_3). The differential surface of the bottom shaded side is

$$d\overline{\sigma}_{bottom} = -dq_1 dq_2 h_1 h_2 \hat{\mathbf{q}}_3 \Big|_{(q_1, q_2, q_3)}, \tag{3.24}$$

where the minus sign occurs because the surface normal is antiparallel to $\hat{\mathbf{q}}_3$. In contrast, the differential surface of the top shaded side is

$$d\overline{\sigma}_{top} = dq_1 dq_2 h_1 h_2 \hat{\mathbf{q}}_3 \Big|_{(q_1, q_2, q_3 + dq_3)}. \tag{3.25}$$

The minus sign is absent because now the surface normal is parallel to $\hat{\mathbf{q}}_3$. In this case, h_1, h_2, and the basis vector $\hat{\mathbf{q}}_3$ are evaluated at the point $(q_1, q_2, q_3 + dq_3)$.

3.4.3 The Displacement Vector

The displacement vector $d\overline{\mathbf{r}}$ plays a central role in the mathematics of curvilinear systems. Once the form of $d\overline{\mathbf{r}}$ is known, equations for most of the vector operations can be easily determined. From multivariable, differential calculus, $d\overline{\mathbf{r}}$ can be written

$$d\overline{\mathbf{r}} = \frac{\partial \overline{\mathbf{r}}}{\partial q_i} dq_i. \tag{3.26}$$

As we showed in Equation 3.22, this can be written using the scale factors as

$$d\overline{\mathbf{r}} = dq_i h_i \hat{\mathbf{q}}_i. \tag{3.27}$$

In a Cartesian system $q_i = x_i$, $\hat{\mathbf{q}}_i = \hat{\mathbf{e}}_i$, and $h_i = 1$, so Equation 3.27 becomes the familiar

$$d\overline{\mathbf{r}} = dx_i \hat{\mathbf{e}}_i. \tag{3.28}$$

In cylindrical coordinates, $h_1 \equiv h_\rho = 1$, $h_2 \equiv h_\phi = \rho$, and $h_3 \equiv h_z = 1$ so

$$d\overline{\mathbf{r}} = dq_1 \hat{\mathbf{q}}_1 + \rho dq_2 \hat{\mathbf{q}}_2 + dq_3 \hat{\mathbf{q}}_3$$
$$= d\rho \hat{\mathbf{e}}_\rho + \rho d\phi \hat{\mathbf{e}}_\phi + dz \hat{\mathbf{e}}_z. \tag{3.29}$$

3.4.4 Vector Products

Because curvilinear systems are orthonormal, we have

$$\hat{\mathbf{q}}_i \cdot \hat{\mathbf{q}}_j = \delta_{ij}. \tag{3.30}$$

This means that the dot product of two vectors, $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$, has the same form as in a Cartesian system:

$$\overline{\mathbf{A}} \cdot \overline{\mathbf{B}} = A_i \hat{\mathbf{q}}_i \cdot B_j \hat{\mathbf{q}}_j = A_i B_j \delta_{ij} = A_i B_i. \tag{3.31}$$

Here A_i and B_i are the curvilinear components of the vectors, which can be obtained by taking axis parallel projections of the vectors onto the basis vectors:

$$A_i \equiv \overline{\mathbf{A}} \cdot \hat{\mathbf{q}}_i. \tag{3.32}$$

With proper ordering, we can always arrange our three curvilinear coordinates to be right-handed. Thus, the form of the cross product is also the same as in a Cartesian system. The cross product of $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ expressed using the Levi-Civita symbol is

$$\overline{\mathbf{A}} \times \overline{\mathbf{B}} = A_i \hat{\mathbf{q}}_i \times B_j \hat{\mathbf{q}}_j = A_i B_j \hat{\mathbf{q}}_k \epsilon_{ijk}. \tag{3.33}$$

3.4.5 The Line Integral

Using the expression for the displacement vector in Equation 3.27, line integrals in curvilinear systems are straightforward:

$$\int_C d\overline{\mathbf{r}} \cdot \overline{\mathbf{V}} = \int_C dq_j h_j \hat{\mathbf{q}}_j \cdot V_i \hat{\mathbf{q}}_i. \tag{3.34}$$

There is a sum over both i and j on the RHS of this equation. Because the curvilinear basis vectors are orthonormal, this line integral becomes

$$\int_C d\overline{\mathbf{r}} \cdot \overline{\mathbf{V}} = \int_C dq_i h_i V_i. \tag{3.35}$$

3.4.6 The Surface Integral

Curvilinear surface integrals are a bit more complicated, because the orientations of the surfaces must be considered. Recalling Figure 3.9, and Equations 3.24 and 3.25, the surface integral of a vector $\overline{\mathbf{V}}$ is

$$\int_S d\overline{\boldsymbol{\sigma}} \cdot \overline{\mathbf{V}} = \int_S \pm dq_1 dq_2 h_1 h_2 V_3 \pm dq_2 dq_3 h_2 h_3 V_1 \pm dq_1 dq_3 h_1 h_3 V_2, \tag{3.36}$$

where each plus or minus sign must be chosen depending on the sign of $d\overline{\boldsymbol{\sigma}} \cdot \hat{\mathbf{q}}_i$.

3.4.7   The Volume Integral

The geometry of Figure 3.9 can be used to derive the form of volume integrals in curvilinear systems. The volume of the element, in the infinitesimal limit, is simply $d\tau = dq_1 dq_2 dq_3 h_1 h_2 h_3$. Therefore the integral of a function $\rho(\bar{r})$ over any volume V is expressed as

$$\int_V d\tau\, \rho(\bar{r}) = \int_V dq_1 dq_2 dq_3 h_1 h_2 h_3 \rho(q_1, q_2, q_3). \tag{3.37}$$

3.4.8   The Gradient

In Chapter 2, we showed how the gradient of a scalar function Φ is defined such that

$$d\Phi = \overline{\nabla}\Phi \cdot d\bar{r}. \tag{3.38}$$

Substitution of Equation 3.27 for $d\bar{r}$ yields

$$d\Phi = \overline{\nabla}\Phi \cdot dq_j h_j \hat{q}_j. \tag{3.39}$$

Differential calculus implies $d\Phi = (\partial\Phi/\partial q_i) dq_i$, so

$$\frac{\partial\Phi}{\partial q_i} dq_i = \overline{\nabla}\Phi \cdot dq_j h_j \hat{q}_j. \tag{3.40}$$

The only way Equation 3.40 can be satisfied is if

$$\overline{\nabla}\Phi = \frac{1}{h_i}\left(\frac{\partial\Phi}{\partial q_i}\right)\hat{q}_i. \tag{3.41}$$

The brackets are necessary in Equation 3.41 to keep the $\partial/\partial q_i$ operator from acting on the basis vectors, because in general $\partial\hat{q}_i/\partial q_j \neq 0$.

3.4.9   The Divergence

The divergence operation in a curvilinear system is more complicated than the gradient, and must be obtained from its integral definition

$$\overline{\nabla} \cdot \mathbf{A} = \lim_{S,V \to 0} \frac{\oint_S d\overline{\sigma} \cdot \mathbf{A}}{\int_V d\tau}, \tag{3.42}$$

where S is the closed surface that encloses the vanishingly small volume V.

Once again, consider the differential volume of Figure 3.9. The denominator of Equation 3.42 for this volume in the infinitesimal limit is straightforward:

$$\int_V d\tau = dq_1 dq_2 dq_3 h_1 h_2 h_3. \tag{3.43}$$

To evaluate the numerator, integration over all six surfaces of V must be performed. First, consider the two shaded sides of Figure 3.9, with normals aligned either parallel or antiparallel to $\hat{q}_3$. The integral over the "bottom" surface is

$$\int_{bottom} d\overline{\sigma} \cdot \overline{A} = -dq_1 dq_2 (h_1 h_2 A_3)\Big|_{(q_1, q_2, q_3)}. \tag{3.44}$$

The minus sign arises because on this surface $d\overline{\sigma}$ and $\hat{q}_3$ are antiparallel. Also notice that A_3, h_1, and h_2 are all functions of the curvilinear coordinates and are evaluated at (q_1, q_2, q_3), the initial values of the coordinates on this surface. The integral over the "top" surface is

$$\int_{top} d\overline{\sigma} \cdot \overline{A} = +dq_1 dq_2 (h_1 h_2 A_3)\Big|_{(q_1, q_2, q_3 + dq_3)}. \tag{3.45}$$

In this case there is no minus sign because this surface normal is oriented parallel to $\hat{q}_3$. The initial value of the q_3 coordinate for this surface has changed by an amount dq_3 as compared to the bottom surface and thus A_3, h_1, and h_2 must all be evaluated at the point $(q_1, q_2, q_3 + dq_3)$. In the differential limit

$$(h_1 h_2 A_3)\Big|_{(q_1, q_2, q_3 + dq_3)} = (h_1 h_2 A_3)\Big|_{(q_1, q_2, q_3)} + \frac{\partial(h_1 h_2 A_3)}{\partial q_3}\Big|_{(q_1, q_2, q_3)} dq_3, \tag{3.46}$$

so the sum of Equations 3.44 and 3.45 is

$$\int_{both} d\overline{\sigma} \cdot \overline{A} = \frac{\partial(h_1 h_2 A_3)}{\partial q_3} dq_1 dq_2 dq_3. \tag{3.47}$$

Combining this result with similar integrations over the remaining four surfaces gives

$$\int_S d\overline{\sigma} \cdot \overline{A} = \left[\frac{\partial(h_2 h_3 A_1)}{\partial q_1} + \frac{\partial(h_1 h_3 A_2)}{\partial q_2} + \frac{\partial(h_1 h_2 A_3)}{\partial q_3} \right] dq_1 dq_2 dq_3. \tag{3.48}$$

Substituting Equations 3.43 and 3.48 into Equation 3.42 gives the result

$$\overline{\nabla} \cdot \overline{A} = \frac{1}{h_1 h_2 h_3} \left[\frac{\partial(h_2 h_3 A_1)}{\partial q_1} + \frac{\partial(h_1 h_3 A_2)}{\partial q_2} + \frac{\partial(h_1 h_2 A_3)}{\partial q_3} \right]. \tag{3.49}$$

3.4.10   The Curl

The curl operation for a curvilinear coordinate system can also be derived from its integral definition:

$$\overline{\nabla} \times \overline{A} \cdot \lim_{S \to 0} \int_S d\overline{\sigma} = \lim_{C \to 0} \oint_C d\overline{r} \cdot \overline{A}, \tag{3.50}$$

where C is a closed path surrounding the surface S, and the direction of $d\overline{\sigma}$ is defined via C and a right-hand convention.
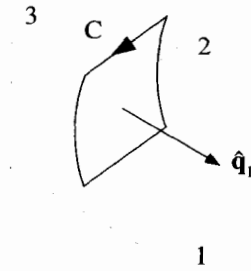
Figure 3.10 Orientation of Surface for Curvilinear Curl Integration

A single component of the curl can be picked out by orienting $d\overline{\sigma}$ along the direction of a basis vector. Consider Figure 3.10, where $d\overline{\sigma}$ is oriented to pick out the $\hat{q}_1$ component. In this case, $d\overline{\sigma} = h_2 dq_2 h_3 dq_3 \hat{q}_1$, so the left side of Equation 3.50 in the differential limit becomes

$$\overline{\nabla} \times \overline{A} \cdot \lim_{S \to 0} \int_S d\overline{\sigma} = dq_2 dq_3 h_2 h_3 \underbrace{\hat{q}_1 \cdot \left[\overline{\nabla} \times \overline{A}\right]}_{[\overline{\nabla} \times \overline{A}]_1}. \tag{3.51}$$

The line integral on the right side of Equation 3.50 naturally divides into four parts along C_a, C_b, C_c, and C_d, as shown in Figure 3.11. The complete integral is then given by

$$\oint_C d\overline{r} \cdot \overline{A} = \int_{C_a} dq_2 h_2 A_2 + \int_{C_b} dq_3 h_3 A_3 + \int_{C_c} dq_2 h_2 A_2 + \int_{C_d} dq_3 h_3 A_3. \tag{3.52}$$

Figure 3.11 Differential Geometry for Curvilinear Curl Integrations

In the differential limit, the integral along C_a evaluates to

$$\int_{C_a} dq_2 h_2 A_2 = (h_2 A_2)\Big|_{(q_1, q_2, q_3)} dq_2, \tag{3.53}$$

where we evaluate both A_2 and h_2 at the point (q_1, q_2, q_3). Likewise, the contribution along C_c is

$$\int_{C_c} dq_2 h_2 A_2 = -(h_2 A_2)\Big|_{(q_1, q_2, q_3 + dq_3)} dq_2, \tag{3.54}$$
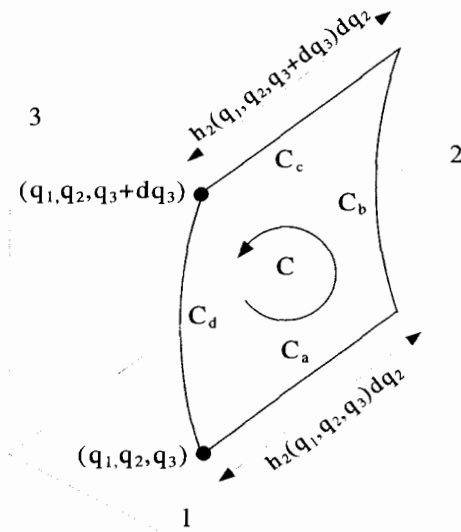
where now we evaluate A_2 and h_2 at $(q_1, q_2, q_3 + dq_3)$. In the differential limit,

$$(h_2 A_2)\Big|_{(q_1, q_2, q_3 + dq_3)} = (h_2 A_2)\Big|_{(q_1, q_2, q_3)} + \frac{\partial(h_2 A_2)}{\partial q_3}\Big|_{(q_1, q_2, q_3)} dq_3, \tag{3.55}$$

which allows the integrals along C_a and C_c to be combined to give

$$\int_{C_a + C_c} d\bar{\mathbf{r}} \cdot \overline{\mathbf{A}} = -\frac{\partial(h_2 A_2)}{\partial q_3}\Big|_{(q_1, q_2, q_3)} dq_2 dq_3. \tag{3.56}$$

Similar integrations can be performed along C_b and C_d. The combination of all four pieces yields

$$\lim_{C \to 0} \oint_C \overline{\mathbf{A}} \cdot d\bar{\mathbf{r}} = \left[\frac{\partial(h_3 A_3)}{\partial q_2} - \frac{\partial(h_2 A_2)}{\partial q_3} \right] dq_2 dq_3. \tag{3.57}$$

Substituting Equations 3.57 and 3.51 into Equation 3.50 gives the 1-component of the curl of $\overline{\mathbf{A}}$:

$$\left[\overline{\mathbf{\nabla}} \times \overline{\mathbf{A}} \right]_1 = \frac{1}{h_2 h_3} \left[\frac{\partial(h_3 A_3)}{\partial q_2} - \frac{\partial(h_2 A_2)}{\partial q_3} \right]. \tag{3.58}$$

The other components of $\overline{\mathbf{\nabla}} \times \overline{\mathbf{A}}$ can be obtained by reorientating the surface shown in Figure 3.10. The results are

$$\left[\overline{\mathbf{\nabla}} \times \overline{\mathbf{A}} \right]_2 = \frac{1}{h_1 h_3} \left[\frac{\partial(h_1 A_1)}{\partial q_3} - \frac{\partial(h_3 A_3)}{\partial q_1} \right] \tag{3.59}$$

$$\left[\overline{\mathbf{\nabla}} \times \overline{\mathbf{A}} \right]_3 = \frac{1}{h_1 h_2} \left[\frac{\partial(h_2 A_2)}{\partial q_1} - \frac{\partial(h_1 A_1)}{\partial q_2} \right]. \tag{3.60}$$

Equations 3.58–3.60 can be written more compactly using a determinant,

$$\overline{\mathbf{\nabla}} \times \overline{\mathbf{A}} = \frac{1}{h_1 h_2 h_3} \begin{vmatrix} h_1 \hat{\mathbf{q}}_1 & h_2 \hat{\mathbf{q}}_2 & h_3 \hat{\mathbf{q}}_3 \\ \partial/\partial q_1 & \partial/\partial q_2 & \partial/\partial q_3 \\ h_1 A_1 & h_2 A_2 & h_3 A_3 \end{vmatrix}_{det}, \tag{3.61}$$

or using the Levi-Civita symbol and subscript notation,

$$\overline{\nabla} \times \overline{\mathbf{A}} = \frac{\epsilon_{ijk}}{h_j h_k} \frac{\partial (h_k A_k)}{\partial q_j} \hat{\mathbf{q}}_i. \tag{3.62}$$

3.5   THE GRADIENT, DIVERGENCE, AND CURL IN CYLINDRICAL AND SPHERICAL SYSTEMS

3.5.1   Cylindrical Operations

In the cylindrical system $h_1 \equiv h_\rho = 1$, $h_2 \equiv h_\phi = \rho$, and $h_3 \equiv h_z = 1$. The gradient, divergence, and curl become

$$\overline{\nabla} \Phi = \frac{\partial \Phi}{\partial \rho} \hat{\mathbf{q}}_\rho + \frac{1}{\rho} \frac{\partial \Phi}{\partial \phi} \hat{\mathbf{q}}_\phi + \frac{\partial \Phi}{\partial z} \hat{\mathbf{q}}_z \tag{3.63}$$

$$\overline{\nabla} \cdot \overline{\mathbf{A}} = \frac{1}{\rho} \frac{\partial (\rho A_\rho)}{\partial \rho} + \frac{1}{\rho} \frac{\partial A_\phi}{\partial \phi} + \frac{\partial A_z}{\partial z} \tag{3.64}$$

$$\overline{\nabla} \times \overline{\mathbf{A}} = \left[\frac{1}{\rho} \frac{\partial A_z}{\partial \phi} - \frac{\partial A_\phi}{\partial z} \right] \hat{\mathbf{q}}_\rho$$

$$+ \left[\frac{\partial A_\rho}{\partial z} - \frac{\partial A_z}{\partial \rho} \right] \hat{\mathbf{q}}_\phi$$

$$+ \frac{1}{\rho} \left[\frac{\partial (\rho A_\phi)}{\partial \rho} - \frac{\partial A_\rho}{\partial \phi} \right] \hat{\mathbf{q}}_z. \tag{3.65}$$

3.5.2   The Spherical Operations

In the spherical system $h_1 \equiv h_r = 1$, $h_2 \equiv h_\theta = r$, and $h_3 \equiv h_\phi = r \sin \theta$. The gradient, divergence, and curl become

$$\overline{\nabla} \Phi = \frac{\partial \Phi}{\partial r} \hat{\mathbf{q}}_r + \frac{1}{r} \frac{\partial \Phi}{\partial \theta} \hat{\mathbf{q}}_\theta + \frac{1}{r \sin \theta} \frac{\partial \Phi}{\partial \phi} \hat{\mathbf{q}}_\phi \tag{3.66}$$

$$\overline{\nabla} \cdot \overline{\mathbf{A}} = \frac{1}{r^2} \frac{\partial (r^2 A_r)}{\partial r} + \frac{1}{r \sin \theta} \frac{\partial (\sin \theta A_\theta)}{\partial \theta} + \frac{1}{r \sin \theta} \frac{\partial A_\phi}{\partial \phi} \tag{3.67}$$

$$\overline{\nabla} \times \overline{\mathbf{A}} = \frac{1}{r \sin \theta} \left[\frac{\partial (\sin \theta A_\phi)}{\partial \theta} - \frac{\partial A_\theta}{\partial \phi} \right] \hat{\mathbf{q}}_r$$

$$+ \frac{1}{r} \left[\frac{1}{\sin \theta} \frac{\partial A_r}{\partial \phi} - \frac{\partial (r A_\phi)}{\partial r} \right] \hat{\mathbf{q}}_\theta$$

$$+ \frac{1}{r} \left[\frac{\partial (r A_\theta)}{\partial r} - \frac{\partial A_r}{\partial \theta} \right] \hat{\mathbf{q}}_\phi. \tag{3.68}$$

EXERCISES FOR CHAPTER 3

1. A vector expressed in a Cartesian system has the following form:

$$\vec{V} = \hat{e}_x + \hat{e}_y + \hat{e}_z.$$

   (a) If this vector exists at the Cartesian point $(1, 2, 1)$, what are its components expressed in a cylindrical system?

   (b) If this vector exists at the Cartesian point $(1, 2, 1)$, what are its components expressed in a spherical system?

   (c) If this vector exists at the Cartesian point $(0, 0, 1)$, what are its components expressed in a cylindrical system?

   (d) If this vector exists at the Cartesian point $(0, 0, 1)$, what are its components expressed in a spherical system?

2. In a cylindrical system, with coordinates (ρ, ϕ, z), the vector $\overline{V}_1$ exists at the point $(1, 0, 0)$ and is given by $\overline{V}_1 = \hat{e}_\rho + \hat{e}_\phi$. A second vector $\overline{V}_2$ exists in the same coordinate system at the point $(1, \pi/2, 0)$ and is given by $\overline{V}_2 = \hat{e}_\rho + \hat{e}_\phi$. What is the value of $\overline{V}_1 \cdot \overline{V}_2$?

3. A circular disk, centered at the origin of a coordinate system, is rotating at a constant angular velocity ω_o.
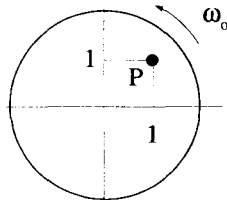
   (a) In Cartesian coordinates, what is the velocity vector of the point P located at the Cartesian coordinates $(1, 1)$?

   (b) In polar coordinates, what is the velocity vector of the point P located at the Cartesian coordinates $(1, 1)$?

4. Perform the integral

$$\int_S d\overline{\sigma} \cdot \overline{F},$$

   where

$$\overline{F} = 2 \sin\theta \, \sin\phi \, \hat{e}_r + 3 \cos\theta \, \cos\phi \, \hat{e}_\phi + 4 \cos\theta \, \cos\phi \, \hat{e}_\theta,$$

and S is one-eighth of a spherical shell of radius 2, centered at the origin of the spherical system. The integration region lies in the positive xyz-octant of a Cartesian system, as shown below:
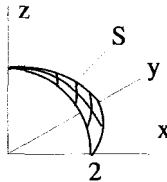
5. Perform the line integral

$$\int_C d\bar{\mathbf{r}} \cdot \hat{\mathbf{e}}_\theta,$$

where $\hat{\mathbf{e}}_\theta$ is a spherical basis vector, and C is the quarter circle of unit radius lying in one quadrant of the Cartesian xz-plane shown below. Evaluate this integral in three ways, using Cartesian, cylindrical, and spherical coordinate systems. What happens to the value of the integral if the direction along C is reversed?
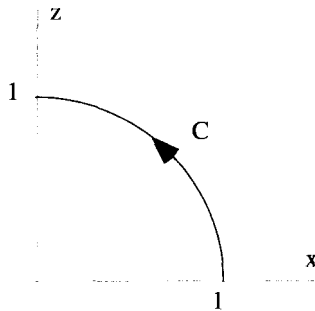
6. Perform the following line integrals along one quarter of a unit radius circle, traversed in a counterclockwise direction. The center of the circle is at the origin of a Cartesian coordinate system, and the quarter circle of interest is in the first xy-quadrant.
(a) $\int_C d\bar{\mathbf{r}} \cdot \bar{\mathbf{V}}$, where $\bar{\mathbf{V}} = 2x\,\hat{\mathbf{e}}_x$.
(b) $\int_C d\bar{\mathbf{r}}\,\Phi$, where $\Phi = 2x$.
(c) Repeat parts (a) and (b) using a polar coordinate system.

7. Determine $\bar{\nabla} \cdot \bar{\mathbf{r}}$, where $\bar{\mathbf{r}}$ is the position vector

$$\bar{\mathbf{r}} = x\,\hat{\mathbf{e}}_x + y\,\hat{\mathbf{e}}_y + z\,\hat{\mathbf{e}}_z$$

and show that $\bar{\nabla} \cdot \bar{\mathbf{r}} = 3$. Repeat this calculation using spherical coordinates.

8. Derive the h_i's for the cylindrical and spherical systems. Also derive the expressions given in the chapter for the gradient, divergence, and curl operators in these systems.

9. Redo Exercise 8 of Chapter 2 using spherical coordinates.

10. A Tokamak fusion device has a geometry that takes the shape of a doughnut or torus. Calculations for such a device are sometimes done with the toroidal coordinates shown in the figure below.
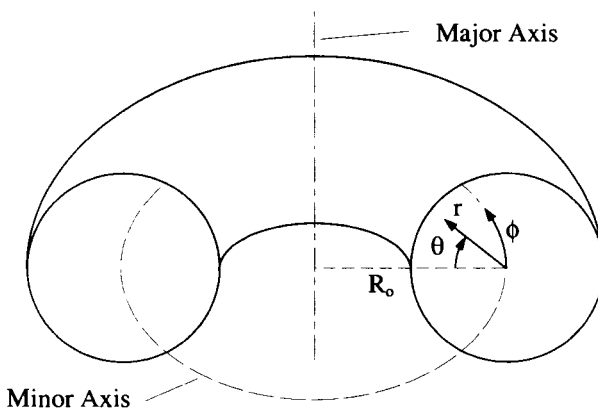
The "major axis" of the device is a vertical, straight line that forms the toroidal axis of symmetry. The "minor axis" is a circle of fixed radius R_o that passes through the center of the doughnut. The position of a point is described by the coordinates (r, θ, ϕ). The coordinates r and θ are similar to a two-dimensional polar system, aligned perpendicular to the minor axis. The coordinate ϕ measures the angular position along the minor axis.

(a) Make a sketch of the unit basis vectors for this toroidal system. Are these vectors orthogonal? Do they form a right-handed system?

(b) Obtain expressions relating the toroidal coordinates to a set of Cartesian coordinates.

(c) Obtain the h_i scale factors for the toroidal system.

(d) Write expressions for the displacement vector $d\bar{r}$, a differential surface area $d\bar{\sigma}$, and a differential volume $d\tau$ in this system.

(e) Write expressions for the gradient $\overline{\nabla}\Phi$, divergence $\overline{\nabla} \cdot \overline{A}$, and curl $\overline{\nabla} \times \overline{A}$ operations in this system.

(f) Laplace's equation written in vector notation is $\nabla^2\Phi = 0$. What is Laplace's equation expressed in these toroidal coordinates?

11. A second toroidal system using coordinates (ξ, η, φ) can be formed, with these coordinates related to the Cartesian coordinates by:

$$x = \frac{a \sinh \eta \cos \varphi}{\cosh \eta - \cos \xi}$$

$$y = \frac{a \sinh \eta \sin \varphi}{\cosh \eta - \cos \xi}$$

$$z = \frac{a \sin \xi}{\cosh \eta - \cos \xi}.$$

(a) Describe the surfaces of constant ξ and constant η.

(b) Pick a point and sketch the basis vectors. Is this a right-handed system?

(c) Obtain the h_i scale factors for this toroidal system.

(d) Express the position and displacement vectors in this system.

12. The (u, v, z) coordinates of an elliptical system are related to a set of Cartesian coordinates by the equations:

$$x = a \cosh u \cos v$$

$$y = a \sinh u \sin v$$

$$z = z.$$

(a) Sketch the lines of constant u and constant v on a two-dimensional xy-grid.

(b) Obtain the h_i scale factors associated with this elliptical system.

(c) Obtain the form of a differential path length, as well as the area and volume elements used to form line, surface, and volume integrals in this elliptical system.

(d) Obtain expressions for the gradient, divergence, and curl in the elliptical system.

(e) Express the position and displacement vectors in the elliptical system.

13. A hyperbolic (u, v, z) coordinate system is sometimes used in electrostatics and hydrodynamics. These coordinates are related to the "standard" Cartesian coordinates by the following equations:

$$2xy = u$$

$$x^2 - y^2 = v$$

$$z = z.$$

(a) Sketch the lines of constant u and constant v in the xy-plane. Note that these lines become surfaces in three dimensions.

(b) Indicate the directions of the basis vectors $\hat{q}_u$ and $\hat{q}_v$ in all four quadrants. Is this system orthogonal?

(c) This (u, v, z) system is left-handed. What can be done to make it right-handed?

(d) In the right-handed version of this system, obtain the h_i scale factors.

(e) Express the position and displacement vectors in this system.

14. An orthogonal (u, v, z) coordinate system is defined by the following set of equations that relate these coordinates to a "standard" Cartesian set:

$$u = \frac{x}{x^2 + y^2}$$

$$v = \frac{-y}{x^2 + y^2}$$

$$z = z.$$

(a) Sketch the lines of constant u and constant v in the xy-plane.

(b) Pick a point in the xy-plane and sketch the $\hat{\mathbf{q}}_u$, $\hat{\mathbf{q}}_v$, and $\hat{\mathbf{q}}_z$ basis vectors. It is not necessary to solve for them in terms of the Cartesian basis vectors $\hat{\mathbf{e}}_x$, $\hat{\mathbf{e}}_y$, and $\hat{\mathbf{e}}_z$.

(c) Is this (u, v, z) system right-handed?

(d) Invert the above expressions and solve for x and y in terms of u and v.

(e) Express the position and displacement vectors in this system.

(f) Express the gradient operation $\overline{\boldsymbol{\nabla}}\Phi(u, v, z)$ in this system.

15. A particle is moving in three-dimensional space. Describe its position $\overline{\mathbf{r}}(t)$, velocity $\dot{\overline{\mathbf{r}}}$, and acceleration $\ddot{\overline{\mathbf{r}}}$ in terms of the the spherical coordinates and basis vectors. To work this problem it will be necessary to take time derivatives of the basis vectors, because they will change as the particle moves in space.

16. A particle is moving in a circular orbit with its position vector given by

$$\overline{\mathbf{r}} = r_o \cos(\omega_o t)\,\hat{\mathbf{e}}_x + r_o \sin(\omega_o t)\,\hat{\mathbf{e}}_y,$$

where r_o and ω_o are constants. If $\dot{\overline{\mathbf{r}}}$ and $\ddot{\overline{\mathbf{r}}}$ are the first and second derivatives of $\overline{\mathbf{r}}$,

(a) Sketch the orbit in the xy-plane.

(b) Evaluate $\overline{\mathbf{r}} \cdot \dot{\overline{\mathbf{r}}}$ and discuss your result.

(c) Show that $\ddot{\overline{\mathbf{r}}} + \omega_o^2 \overline{\mathbf{r}} = 0$.

17. Consider a scalar function Φ of position which depends on the cylindrical coordinates (ρ, ϕ, z) such that $\Phi = \rho \cos \phi$.

(a) Plot the surfaces of constant Φ in the xy-plane.

(b) Let the vector $\overline{\mathbf{V}} = \overline{\boldsymbol{\nabla}}\Phi$, and find the radial V_ρ and azimuthal V_ϕ components of $\overline{\mathbf{V}}$.

(c) The field lines for $\overline{V}$ can be described by a function $\rho = \rho(\phi)$. Show that $\rho(\phi)$ satisfies the equation

$$\frac{d\rho(\phi)}{d\phi} = \frac{\rho V_\rho}{V_\phi}.$$

(d) Solve the above equation for $\rho(\phi)$, and plot the $\overline{V}$ field lines on the same graph you plotted the surfaces of constant Φ.

18. A flat disk rotates about an axis normal to its plane and passing through its center. Show that the velocity vector $\overline{v}$ of any point on the disk satisfies the equations

$$\overline{\nabla} \cdot \overline{v} = 0$$
$$\overline{\nabla} \times \overline{v} = 2\overline{\omega},$$

where $\overline{\omega}$ is the angular velocity vector of the disk. This vector is defined as

$$\overline{\omega} = \frac{\overline{r} \times \overline{v}}{r^2}.$$

19. Consider a sphere of radius r_o, rotating at a constant angular rate ω_o about its z-axis so that

$$\overline{\omega} = \omega_o \hat{e}_z.$$

(a) Find an expression for the velocity vector $\overline{v}$ of points on the surface of the sphere using spherical coordinates and spherical basis vectors. Remember $\overline{v} = \overline{\omega} \times \overline{r}$.

(b) Perform the integration

$$\oint d\overline{r} \cdot \overline{v}$$

around the equator of the sphere.

(c) Now perform the surface integral

$$\oint d\overline{\sigma} \times \overline{v}$$

over the entire surface of the sphere.

20. The magnetic field inside an infinitely long solenoid is uniform

$$\overline{B} = B_o \hat{e}_z.$$

Determine the vector potential $\overline{A}$ such that $\overline{\nabla} \times \overline{A} = \overline{B}$. For this situation, the magnetic field can also be derived from a scalar potential, $\overline{B} = -\overline{\nabla}\Phi$. Why is this the case and what is Φ?

The magnetic field inside a straight wire, aligned with the z-axis and carrying a uniformly distributed current, can be expressed using cylindrical coordinates as

$$\overline{\mathbf{B}} = \frac{B_o \rho}{\rho_o} \hat{\mathbf{e}}_\phi,$$

where B_o and ρ_o are constants. In this case the magnetic field can still be obtained from a vector potential $\overline{\nabla} \times \overline{\mathbf{A}} = \overline{\mathbf{B}}$. Find this $\overline{\mathbf{A}}$. Now, however, it is no longer possible to find a scalar potential such that $\overline{\mathbf{B}} = -\overline{\nabla}\Phi$. Why is this the case?

21. A static magnetic field is related to the current density by one of Maxwell's equations,

$$\overline{\nabla} \times \overline{\mathbf{B}} = \mu_o \overline{\mathbf{J}}.$$

(a) If the magnetic field for $\rho < \rho_o$ is given, in cylindrical coordinates, by

$$\overline{\mathbf{B}} = \frac{B_o \rho^3}{\rho_o^3} \hat{\mathbf{e}}_\phi,$$

how is the current density $\overline{\mathbf{J}}$ distributed?

(b) If $\overline{\mathbf{B}} = 0$ for $\rho > \rho_o$, how much current must be flowing in a cylindrical shell at $\rho = \rho_o$?

22. Classically, the angular momentum $\overline{\mathbf{L}}$ is given by

$$\overline{\mathbf{L}} = \overline{\mathbf{r}} \times \overline{\mathbf{p}},$$

where $\overline{\mathbf{r}}$ is the position vector of an object and $\overline{\mathbf{p}}$ is its linear momentum.

In quantum mechanics, the value of the angular momentum is obtained by letting a vector differential operator $\overline{\mathcal{L}}_{op}$ act on a wave function Ψ:

$$\text{Angular Momentum} = \overline{\mathcal{L}}_{op} \Psi.$$

The angular momentum operator can be obtained using the correspondence principle of quantum mechanics. This principle says that the operators are obtained by replacing classical objects in the classical equations with operators. Therefore,

$$\overline{\mathcal{L}}_{op} = \overline{\mathcal{R}}_{op} \times \overline{\mathcal{P}}_{op}.$$

The position operator is just the position vector

$$\overline{\mathcal{R}}_{op} = \overline{\mathbf{r}}.$$

The momentum operator is

$$\overline{\mathcal{P}}_{op} = -i\hbar\overline{\nabla},$$

where i is the square root of -1 and $\hbar$ is Planck's constant divided by 2π. Show that, in Cartesian geometry,

$$\overline{\mathcal{L}}_{op} = -i\hbar \left[y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right] \hat{\mathbf{e}}_x$$

$$-i\hbar \left[z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right] \hat{\mathbf{e}}_y$$

$$-i\hbar \left[x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right] \hat{\mathbf{e}}_z.$$

23. The derivation of the curvilinear divergence in this chapter is somewhat sloppy, because it treats the scale factors as constants along each surface of the volume. Derive the same expression for the divergence following the more rigorous method demonstrated in Appendix B for the curvilinear curl.

4

INTRODUCTION TO TENSORS

Tensors pervade all the branches of physics. In mechanics, the moment of inertia tensor is used to describe the rotation of rigid bodies, and the stress-strain tensor describes the deformation of elastic bodies. In electromagnetism, the conductivity tensor extends Ohm's law to handle current flow in nonisotropic media, and Maxwell's stress tensor is the most elegant way to deal with electromagnetic forces. The metric tensor of relativistic mechanics describes the strange geometry of space and time.

This chapter presents an introduction to tensors and their manipulations, first strictly in Cartesian coordinates and then in generalized curvilinear coordinates. We will limit the material in this chapter to orthonormal coordinate systems. Tensors in nonorthonormal systems are discussed later, in Chapter 14. At the end of the chapter, we introduce the so-called "pseudo"-objects, which arise when we consider transformations between right- and left-handed coordinate systems.

4.1 THE CONDUCTIVITY TENSOR AND OHM'S LAW

The need for tensors can easily be demonstrated by considering Ohm's law. In an ideal resistor, Ohm's law relates the current to the voltage in the linear expression

$$I = \frac{V}{R}. \tag{4.1}$$

In this equation, I is the current through the resistor, and V is the voltage applied across it. Using MKS units, I is measured in amperes, V in volts, and R in ohms.

Equation 4.1 describes the current flow through a discrete element. To apply Ohm's law to a distributed medium, such as a crystalline solid, an alternative form of this equation is used:

$$\bar{\mathbf{J}} = \sigma \bar{\mathbf{E}}. \tag{4.2}$$

Here $\bar{\mathbf{J}}$ is the current density, $\bar{\mathbf{E}}$ is the electric field, and σ is the material's *conductivity*. In MKS units, $\bar{\mathbf{J}}$ is measured in amperes per meter squared, $\bar{\mathbf{E}}$ in volts per meter, and σ in inverse ohm-meters.

Equation 4.2 describes a very simple physical relationship between the current density and the electric field, because the conductivity has been expressed as a scalar. With a scalar conductivity, the amount of current flow is governed solely by the magnitudes of σ and $\bar{\mathbf{E}}$, while the direction of the flow is always parallel to $\bar{\mathbf{E}}$. But in some materials, this is not always the case. Many crystalline solids allow current to flow more easily in one direction than another. These *nonisotropic* materials must have different conductivities in different directions. In addition, these crystals can even experience current flow perpendicular to an applied electric field. Clearly Equation 4.2, with a scalar conductivity, will not handle these situations.

One solution is to construct an array of conductivity elements and express Ohm's law using matrix notation as

$$\begin{bmatrix} J_1 \\ J_2 \\ J_3 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix}. \tag{4.3}$$

In Equation 4.3, the current density and electric field vectors are represented by column matrices and the conductivity is now a square matrix. This equation can be written in more compact matrix notation as

$$[J] = [\sigma][E] \tag{4.4}$$

or in subscript/summation notation as

$$J_i = \sigma_{ij} E_j. \tag{4.5}$$

All these expressions produce the desired result. Any linear relationship between $\bar{\mathbf{J}}$ and $\bar{\mathbf{E}}$ can be described. The 1-component of the current density is related to the 1-component of the electric field via σ_{11}, while the 2-component of the current density is related to the 2-component of the electric field through σ_{22}. Perpendicular flow is described by the off-diagonal elements. For example, the σ_{12} element describes flow in the 1-direction due to an applied field in the 2-direction.

The matrix representation for nonisotropic conductivity does, however, have a fundamental problem. The elements of the matrix obviously must depend on our choice of coordinate system. Just as with the components of a vector, if we reorient our coordinate system, the specific values in the matrix array must change. The matrix array itself, unfortunately, carries no identification of the coordinate system used. The way we solved this problem for vector quantities was to incorporate the basis vectors

directly into the notation. That same approach can be used to improve the notation for the nonisotropic conductivity. We define a new object, called the conductivity tensor, which we write as $\overline{\overline{\sigma}}$. This tensor includes both the elements of the conductivity matrix array and the basis vectors of the coordinate system in which these elements are valid. Because this notation is motivated by the similar notation we have been using for vectors, we begin with a quick review.

Remember that a vector quantity, such as the electric field, can be represented by a column matrix:

$$\overline{E} \rightarrow \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix}. \tag{4.6}$$

The vector and matrix quantities are not equal, however, because the matrix cannot replace $\overline{E}$ in a vector equation and vice versa. Instead, the basis vectors of the coordinate system in which the vector is expressed must be included to form an equivalent expression:

$$\overline{E} = E_i \hat{e}_i. \tag{4.7}$$

The nonisotropic conductivity tensor $\overline{\overline{\sigma}}$ can be treated in a similar way. It can be represented by the matrix array

$$\overline{\overline{\sigma}} \rightarrow \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}, \tag{4.8}$$

but the matrix array and the tensor are not equivalent because the matrix array contains no information about the coordinate system. Following the pattern used for vectors and the expression for a vector given in Equation 4.7, we express the conductivity tensor as

$$\overline{\overline{\sigma}} = \sigma_{ij} \hat{e}_i \hat{e}_j. \tag{4.9}$$

The discussion that follows will show that this is a very powerful notation. It supports all the algebraic manipulations that the matrix array notation does and also handles the transformations between coordinate systems with ease.

The expression for the conductivity tensor on the RHS of Equation 4.9 contains the elements of the conductivity matrix array and two basis vectors from the coordinate system in which the elements are valid. There is no operation between these basis vectors. They serve as "bins" into which the σ_{ij} elements are placed. There is a double sum over the subscripts i and j and so, for a three-dimensional system, there will be 9 terms in this sum, each containing two of the basis vectors. In other words, we can expand the conductivity as

$$\overline{\overline{\sigma}} = \sum_i \sum_j \sigma_{ij} \hat{e}_i \hat{e}_j = \sigma_{11} \hat{e}_1 \hat{e}_1 + \sigma_{12} \hat{e}_1 \hat{e}_2 + \sigma_{21} \hat{e}_2 \hat{e}_1 + \cdots. \tag{4.10}$$

This is analogous to how a vector can be expanded in terms of its basis vectors as

$$\overline{V} = \sum_i V_i \hat{e}_i = V_1 \hat{e}_1 + V_2 \hat{e}_2 + V_3 \hat{e}_3. \tag{4.11}$$

Let's see how this new notation handles Ohm's law. Using the conductivity tensor, we can write it in coordinate-independent "vector/tensor" notation as

$$\overline{J} = \overline{\overline{\sigma}} \cdot \overline{E}. \tag{4.12}$$

Notice the dot product between the conductivity tensor and the electric field vector on the RHS of this expression. We can write this out in subscript/summation notation as

$$J_s \hat{e}_s = (\sigma_{jk} \hat{e}_j \hat{e}_k) \cdot (E_l \hat{e}_l). \tag{4.13}$$

By convention, the dot product in Equation 4.13 operates between the second basis vector of $\overline{\overline{\sigma}}$ and the single basis vector of $\overline{E}$. We can manipulate Equation 4.13 as follows:

$$J_s \hat{e}_s = \sigma_{jk} E_l \hat{e}_j \hat{e}_k \cdot \hat{e}_l \tag{4.14}$$

$$J_s \hat{e}_s = \sigma_{jk} E_l \hat{e}_j \delta_{kl} \tag{4.15}$$

$$J_s \hat{e}_s = \sigma_{jk} E_k \hat{e}_j. \tag{4.16}$$

The quantities on the left- and right-hand sides of Equation 4.16 are vectors. The ith components of these vectors can be obtained by dot multiplying both sides of Equation 4.16 by $\hat{e}_i$ to give

$$J_i = \sigma_{ik} E_k, \tag{4.17}$$

which is identical to Equations 4.3–4.5. Keep in mind that there is a difference between $\overline{\overline{\sigma}} \cdot \overline{E}$ and $\overline{E} \cdot \overline{\overline{\sigma}}$. The order of the terms matters because, in general,

$$\hat{e}_j \hat{e}_k \cdot \hat{e}_l \neq \hat{e}_l \cdot \hat{e}_j \hat{e}_k. \tag{4.18}$$

The basis vectors in this tensor notation serve several functions:

1. They establish bins to separate the tensor components.
2. They couple the components to a coordinate system.
3. They set up the formalism for tensor algebra operations.
4. As shown later in the chapter, they also simplify the formalism for transformations between coordinate systems.

Now that we have motivated our investigation of tensors with a specific example, we proceed to look at some of their more formal properties.

4.2 GENERAL TENSOR NOTATION AND TERMINOLOGY

The conductivity tensor is a specific example of a tensor that uses two basis vectors and whose elements have two subscripts. In general, a tensor can have any number of subscripts, but the number of subscripts must always be equal to the number of basis vectors. So in general,

$$\overline{\overline{\mathbf{T}}} = T_{ijk...}\,\hat{\mathbf{e}}_i\hat{\mathbf{e}}_j\hat{\mathbf{e}}_k\ldots \tag{4.19}$$

The number of basis vectors determines the *rank* of the tensor. Notice how the tensor notation is actually a generalization of the vector notation used in previous chapters. Vectors are simply tensors of rank one. Scalars can be considered tensors of rank zero. Keep in mind that the rank of the tensor and the dimension of the coordinate system are different quantities. The rank of the tensor identifies the number of basis vectors on the right-hand side of Equation 4.19, while the dimension of the coordinate system determines the number of different values a particular subscript can take. For a three-dimensional system, the subscripts $(i, j, k$, etc.) can each take on the values $(1, 2, 3)$.

This notation introduces the possibility of a new operation between vectors called the *dyadic product*. This product is either written as $\overline{\mathbf{A}}{:}\overline{\mathbf{B}}$ or just $\overline{\mathbf{A}}\,\overline{\mathbf{B}}$. The dyadic product between vectors creates a second-rank tensor,

$$\overline{\mathbf{A}}\,\overline{\mathbf{B}} = A_i\hat{\mathbf{e}}_iB_j\hat{\mathbf{e}}_j = A_iB_j\hat{\mathbf{e}}_i\hat{\mathbf{e}}_j. \tag{4.20}$$

This type of operation can be extended to combine any two tensors of arbitrary rank. The result is a tensor with rank equal to the sum of the ranks of the two tensors in the product. Sometimes this operation is referred to as an outer product, as opposed to the dot product, which is often called an inner product.

4.3 TRANSFORMATIONS BETWEEN COORDINATE SYSTEMS

The new tensor notation of Equation 4.19 makes it easy to transform tensors between different coordinate systems. In fact, many texts formally define a tensor as "an object that transforms as a tensor." While this appears to be a meaningless statement, as will be shown in this section, it is right to the point.

In this chapter, only transformations between orthonormal systems are considered. First, transformations between Cartesian systems are examined and then the results are generalized to curvilinear systems. The complications of transformations in nonorthonormal systems are deferred until Chapter 14.

4.3.1 Vector Transformations Between Cartesian Systems

We begin by looking at vector component transformations between two simple, two-dimensional Cartesian systems. A primed system is rotated by an angle θ_o with
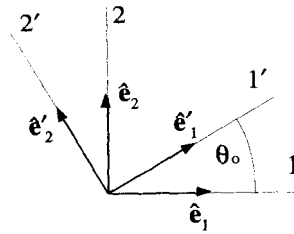
Figure 4.1 Rotated Systems

respect to an unprimed system, as shown in Figure 4.1. A vector $\overline{V}$ can be expressed with unprimed or primed components as

$$\overline{V} = V_i \hat{e}_i = V_i' \hat{e}_i'. \tag{4.21}$$

From the geometry of Figure 4.2, it can be seen that the vector components in the primed system are related to the vector components in the unprimed system by the set of equations:

$$V_1' = \cos \theta_o V_1 + \sin \theta_o V_2$$
$$V_2' = - \sin \theta_o V_1 + \cos \theta_o V_2. \tag{4.22}$$

These equations can be written in matrix notation as

$$[V'] = [a][V], \tag{4.23}$$

where $[V']$ and $[V]$ are column matrices representing the vector $\overline{V}$ with primed and unprimed components, and $[a]$ is the square matrix:

$$[a] = \begin{bmatrix} \cos \theta_o & \sin \theta_o \\ - \sin \theta_o & \cos \theta_o \end{bmatrix} . \tag{4.24}$$
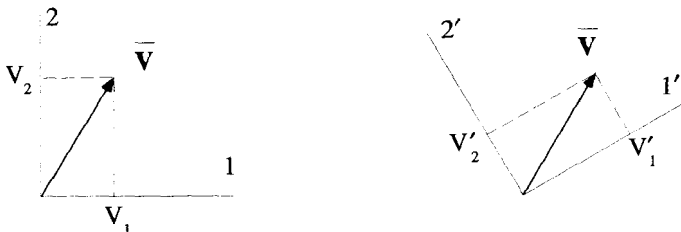

Figure 4.2 Vector Components

4.3.2   The Transformation Matrix

In general, any linear coordinate transformation of a vector can be written, using subscript notation, as

$$V_i' = a_{ij}V_j, \tag{4.25}$$

where $[a]$ is called the transformation matrix. In the discussion that follows, two simple approaches for determining the elements of $[a]$ are presented. The first assumes two systems with known basis vectors. The second assumes knowledge of only the coordinate equations relating the two systems. The choice between methods is simply convenience. While Cartesian coordinate systems are assumed in the derivations, these methods easily generalize to all coordinate transformations.

Determining [a] *from the Basis Vectors*   If the basis vectors of both coordinate systems are known, it is quite simple to determine the components of $[a]$. Consider a vector $\overline{V}$ expressed with components in two different Cartesian systems,

$$\overline{V} = V_k\hat{e}_k = V_i'\hat{e}_i'. \tag{4.26}$$

Substitute the expression for V_i' given in Equation 4.25 into Equation 4.26 to obtain

$$V_k\hat{e}_k = a_{ij}V_j\hat{e}_i'. \tag{4.27}$$

This must be true for any $\overline{V}$. In particular, let $\overline{V} = \hat{e}_m$, one of the unprimed basis vectors (in other words, $V_{k \neq m} = 0$ and $V_{k=m} = 1$), to obtain

$$\hat{e}_m = a_{im}\hat{e}_i'. \tag{4.28}$$

Dot multiplication of $\hat{e}_n'$ on both sides yields

$$a_{nm} = (\hat{e}_n' \cdot \hat{e}_m). \tag{4.29}$$

Notice the elements of $[a]$ are just the directional cosines between all the pairs of basis vectors in the primed and unprimed systems.

Determining [a] *from the Coordinate Equations*   If the basis vectors are not explicitly known, the coordinate equations relating the two systems provide the quickest method for determining the transformation matrix. Begin by considering the expressions for the displacement vector in the two systems. Because both the primed and unprimed systems are Cartesian,

$$d\overline{r} = dx_i\hat{e}_i = dx_i'\hat{e}_i', \tag{4.30}$$

where the dx_i' and dx_i are the total differentials of the coordinates. Because Equation 4.25 holds for the components of any vector, including the displacement vector,

$$dx_i' = a_{ij}dx_j. \tag{4.31}$$

Equation 4.31 provides a general method for obtaining the elements of the matrix $[a]$ using the equations relating the primed and unprimed coordinates. Working in three dimensions, assume these equations are

$$x_1' = x_1'(x_1, x_2, x_3)$$
$$x_2' = x_2'(x_1, x_2, x_3) \qquad (4.32)$$
$$x_3' = x_3'(x_1, x_2, x_3),$$

or more compactly,

$$x_i' = x_i'(x_1, x_2, x_3). \qquad (4.33)$$

Expanding the total differentials of Equations 4.32 gives

$$dx_1' = \frac{\partial x_1'(x_1, x_2, x_3)}{\partial x_1}dx_1 + \frac{\partial x_1'(x_1, x_2, x_3)}{\partial x_2}dx_2 + \frac{\partial x_1'(x_1, x_2, x_3)}{\partial x_3}dx_3$$

$$dx_2' = \frac{\partial x_2'(x_1, x_2, x_3)}{\partial x_1}dx_1 + \frac{\partial x_2'(x_1, x_2, x_3)}{\partial x_2}dx_2 + \frac{\partial x_2'(x_1, x_2, x_3)}{\partial x_3}dx_3$$

$$dx_3' = \frac{\partial x_3'(x_1, x_2, x_3)}{\partial x_1}dx_1 + \frac{\partial x_3'(x_1, x_2, x_3)}{\partial x_2}dx_2 + \frac{\partial x_3'(x_1, x_2, x_3)}{\partial x_3}dx_3.$$

Again, using subscript notation, this is written more succinctly as

$$dx_i' = \frac{\partial x_i'(x_1, x_2, x_3)}{\partial x_j}dx_j. \qquad (4.34)$$

Comparison of Equation 4.31 and Equation 4.34 identifies the elements of $[a]$ as

$$a_{ij} = \frac{\partial x_i'(x_1, x_2, x_3)}{\partial x_j}. \qquad (4.35)$$

The Orthonormal Property of **[a]**   If the original and the primed coordinate systems are both orthonormal, a useful relationship exists among the elements of $[a]$. It is easily derived by dot multiplying both sides of Equation 4.28 by $\hat{e}_k$:

$$\hat{e}_j = a_{ij}\hat{e}_i'$$
$$\hat{e}_j \cdot \hat{e}_k = a_{ij}(\hat{e}_i' \cdot \hat{e}_k) \qquad (4.36)$$
$$\delta_{jk} = a_{ij}a_{ik}.$$

Equation 4.36 written in matrix form is

$$[a][a]^\dagger = [1], \qquad (4.37)$$

where $[a]^\dagger$ is the standard notation for the transpose of $[a]$, and the matrix $[1]$ is a square matrix, with 1's on the diagonal and 0's on the off-diagonal elements.

The Inverse of [a]   The matrix [a] generates primed vector components from unprimed components, as indicated in Equation 4.25. This expression can be inverted with the inverse of [a], which is written as $[a]^{-1}$ and defined by

$$[a][a]^{-1} = [a]^{-1}[a] = [1], \tag{4.38}$$

or in subscript notation

$$a_{ij}^{-1} a_{jk} = a_{ij} a_{jk}^{-1} = \delta_{ik}. \tag{4.39}$$

The subscript notation handles the inversion easily:

$$\begin{aligned}
V_i' &= a_{ij} V_j \\
a_{ki}^{-1} V_i' &= a_{ki}^{-1} a_{ij} V_j \\
a_{ki}^{-1} V_i' &= \delta_{kj} V_j \\
a_{ki}^{-1} V_i' &= V_k.
\end{aligned} \tag{4.40}$$

Transformation matrices which obey the orthonormality condition are simple to invert. Comparison of Equation 4.37 and Equation 4.38 shows that

$$[a]^{-1} = [a]^\dagger, \tag{4.41}$$

or in subscript notation,

$$a_{ij}^{-1} = a_{ji}. \tag{4.42}$$

The inverse relation then becomes

$$V_i = a_{ji} V_j'. \tag{4.43}$$

Basis Vector Transformations   The unprimed basis vectors were related to the primed basis vectors by Equation 4.28:

$$\hat{\mathbf{e}}_i = a_{ji} \hat{\mathbf{e}}_j'. \tag{4.44}$$

Using the fact that the inverse of the [a] matrix is its transpose, this expression can be inverted to give the primed basis vectors in terms of the unprimed ones

$$\hat{\mathbf{e}}_j' = a_{ij} \hat{\mathbf{e}}_i. \tag{4.45}$$

Remember that these expressions are only valid for transformations if both the primed and unprimed systems are orthonormal.

4.3.3   Coordinate Transformation Summary

The box below summarizes the transformation equations between two Cartesian coordinate systems:

$$
\begin{array}{ll}
V_i' = a_{ij}V_j & \hat{\mathbf{e}}_i' = a_{ij}\hat{\mathbf{e}}_j \\[2mm]
V_i = a_{ji}V_j' & \hat{\mathbf{e}}_i = a_{ji}\hat{\mathbf{e}}_j' \\[2mm]
a_{ij} = (\hat{\mathbf{e}}_i' \cdot \hat{\mathbf{e}}_j) = \partial x_i'(x_1, x_2, x_3)/\partial x_j &
\end{array}
$$

The functions $x_i' = x_i'(x_1, x_2, x_3)$ relate the primed Cartesian coordinates to the unprimed Cartesian coordinates. To help keep things straight, notice there is a pattern to these transformation equations. Every time we convert from the unprimed system to the primed system, whether we are dealing with a basis vector or the components of some vector, we sum over the *second* subscript of a_{ij}. In contrast, conversions from the primed system to the unprimed system always sum over the *first* subscript.

4.3.4 Tensor Transformations

To understand why the elements of a tensor must change values when expressed in different coordinate systems, consider the conductivity tensor. If in one set of coordinates, current flows more easily in the 1-direction than the 2-direction, then $\sigma_{11} > \sigma_{22}$. Observation of the same physical situation in a new, primed coordinate system where the 1'-direction is equivalent to the original 2-direction and the 2'-direction is the same as the original 1-direction, would show that $\sigma_{11}' < \sigma_{22}'$. Clearly the elements of the conductivity tensor must take on different values in the two systems, even though they are describing the same physical situation. This is also true of a vector quantity; the same velocity vector will have different components in different coordinate systems.

Tensor transformations follow the same pattern as vector transformations. A vector expressed in a primed and unprimed system is still the same vector:

$$
\overline{\mathbf{V}} = V_i\hat{\mathbf{e}}_i = V_j'\hat{\mathbf{e}}_j'. \tag{4.46}
$$

Likewise, using the notation of Equation 4.19, the expressions for a second-rank tensor in the two systems must obey

$$
\overline{\overline{\mathbf{T}}} = T_{ij}\hat{\mathbf{e}}_i\hat{\mathbf{e}}_j = T_{rs}'\hat{\mathbf{e}}_r'\hat{\mathbf{e}}_s'. \tag{4.47}
$$

Herein lies the beauty of this notation. The relationship between the elements, T_{ij} and T_{rs}', is built into Equation 4.47 and is easily obtained by applying two successive dot products to both sides. The first dot product yields

$$
\begin{aligned}
\hat{\mathbf{e}}_l \cdot T_{ij}\hat{\mathbf{e}}_i\hat{\mathbf{e}}_j &= \hat{\mathbf{e}}_l \cdot T_{rs}'\hat{\mathbf{e}}_r'\hat{\mathbf{e}}_s' \\
T_{ij}(\hat{\mathbf{e}}_l \cdot \hat{\mathbf{e}}_i)\hat{\mathbf{e}}_j &= T_{rs}'(\hat{\mathbf{e}}_l \cdot \hat{\mathbf{e}}_r')\hat{\mathbf{e}}_s' \\
T_{ij}\delta_{li}\hat{\mathbf{e}}_j &= T_{rs}'a_{rl}\hat{\mathbf{e}}_s' \\
T_{lj}\hat{\mathbf{e}}_j &= T_{rs}'a_{rl}\hat{\mathbf{e}}_s'.
\end{aligned} \tag{4.48}
$$

Applying a second dot product in the same manner yields

$$T_{lm} = T'_{rs}a_{rl}a_{sm}.$$ (4.49)

To invert Equation 4.49 use the inverted matrix $[a]^{-1}$ twice and remember, for orthonormal coordinate systems, $a_{ij}^{-1} = a_{ji}$. This gives

$$T'_{lm} = T_{rs}a_{lr}a_{ms}.$$ (4.50)

In general, tensor transformations require one a_{ij} factor for each subscript in the tensor. In other words, an rth rank tensor needs r different a_{ij} factors. If the transformation goes from the unprimed to the primed system, all the a_{ij}'s are summed over their second subscript. For the inverse transformation, going from the primed to the unprimed system, the sums are over the first subscript. Tensor transformations, for arbitrary rank tensors, can be summarized as follows:

$$\boxed{\begin{aligned} T'_{ijk...} &= T_{rst...}a_{ir}a_{js}a_{kt}\cdots \\ T_{ijk...} &= T'_{rst...}a_{ri}a_{sj}a_{tk}\cdots \end{aligned}}$$

where the elements of the matrix $[a]$ are given by Equation 4.35.

There is another important feature in the tensor notation of Equation 4.19. Unlike in a matrix equation where all terms must be in the same basis, the tensor/vector notation allows equations to have mixed bases. Imagine the elements of Ohm's law expressed in both a primed and unprimed coordinate system:

$$\begin{aligned} \bar{\mathbf{J}} &= J_i\hat{\mathbf{e}}_i = J'_i\hat{\mathbf{e}}'_i \\ \bar{\mathbf{E}} &= E_i\hat{\mathbf{e}}_i = E'_i\hat{\mathbf{e}}'_i \\ \bar{\bar{\boldsymbol{\sigma}}} &= \sigma_{ij}\hat{\mathbf{e}}_i\hat{\mathbf{e}}_j = \sigma'_{ij}\hat{\mathbf{e}}'_i\hat{\mathbf{e}}'_j. \end{aligned}$$ (4.51)

Ohm's law reads

$$\bar{\mathbf{J}} = \bar{\bar{\boldsymbol{\sigma}}} \cdot \bar{\mathbf{E}},$$ (4.52)

and any combination of the representations in Equations 4.51 can be used in the evaluation. For example:

$$J_i\hat{\mathbf{e}}_i = (\sigma'_{jk}\hat{\mathbf{e}}'_j\hat{\mathbf{e}}'_k) \cdot (E_l\hat{\mathbf{e}}_l) = \sigma'_{jk}E_l\hat{\mathbf{e}}'_j(\hat{\mathbf{e}}'_k \cdot \hat{\mathbf{e}}_l) = \sigma'_{jk}E_l\hat{\mathbf{e}}'_ja_{kl}.$$ (4.53)

The fact that elements of $\bar{\bar{\boldsymbol{\sigma}}}$ from the primed frame are combined with components of $\bar{\mathbf{E}}$ from the unprimed frame presents no problem. The dot product of the basis vectors takes care of the mixed representations, as long as the order of the basis vectors is preserved. This is accomplished in Equation 4.53 by the fact that $\hat{\mathbf{e}}'_k \cdot \hat{\mathbf{e}}_l \neq \delta_{kl}$. This type of an operation could not be performed using a matrix representation without explicitly converting everything to the same basis.

The value of expressing a tensor in the form of 4.19 should now be clear. In addition to handling the same algebraic manipulations as a matrix array, it also contains all

the information necessary to transform the elements from one coordinate system to another. Thus a tensor is truly a coordinate-independent, geometric object, just as a vector is.

4.4 TENSOR DIAGONALIZATION

In physics and engineering problems we often want to *diagonalize* a tensor. What this means is we want to find a particular coordinate system in which the matrix array representation of a tensor has nonzero elements only along its diagonal. A rigid body will experience no vibration when rotated around any of the three axes of a coordinate system in which the moment of inertia tensor is diagonalized. The process of balancing a wheel of an automobile makes use of this fact. Small, asymmetrically placed weights are added to the rim until one of these special axes lies along the axle.

Many students get lost in the mathematical process of diagonalization and forget that it is actually a transformation of coordinates. In this section, we derive the elements of the transformation matrix $[a]$ that diagonalizes a given tensor. We start off with a completely theoretical treatment of the subject. Then two numerical examples, one nondegenerate and one degenerate, are worked out in detail.

4.4.1 Diagonalization and the Eigenvalue Problem

Based on the discussion of the previous section, a tensor $\overline{\overline{\sigma}}$ written in an unprimed system must be equivalent to the same tensor written in a primed system:

$$\overline{\overline{\sigma}} = \sigma_{ij}\hat{e}_i\hat{e}_j = \sigma'_{st}\hat{e}'_s\hat{e}'_t. \tag{4.54}$$

We are interested in a very special primed system, a system where all the off-diagonal elements of $\overline{\overline{\sigma}}$ are zero. In this case, Equation 4.54 becomes

$$\overline{\overline{\sigma}} = \sigma_{ij}\hat{e}_i\hat{e}_j = \sigma'_{ss}\hat{e}'_s\hat{e}'_s. \tag{4.55}$$

Both the tensor elements and the basis vectors of the unprimed system are presumed to be known. The problem is to find σ'_{ss}, the elements of the tensor in the primed system, and $\hat{e}'_s$, the primed basis vectors, such that Equation 4.55 is satisfied. To do this, form the dot product of Equation 4.55 with the first primed basis vector $\hat{e}'_1$ as follows:

$$\begin{aligned}
\overline{\overline{\sigma}} \cdot \hat{e}'_1 &= \sigma'_{ss}\hat{e}'_s\hat{e}'_s \cdot \hat{e}'_1 \\
&= \sigma'_{ss}\hat{e}'_s\delta_{s1} \\
&= \sigma'_{11}\hat{e}'_1.
\end{aligned} \tag{4.56}$$

Equation 4.56 reveals an important property of the basis vectors of the system where the tensor is diagonal. They do not change direction when dot multiplied by the tensor. They can, however, change in magnitude. If we define $\lambda_1 = \sigma'_{11}$, Equation 4.56

becomes

$$\overline{\overline{\sigma}} \cdot \hat{e}_1' = \lambda_1 \hat{e}_1'. \tag{4.57}$$

The λ_1 factor is called an *eigenvalue* of $\overline{\overline{\sigma}}$. An eigenvalue results when an operation on an object produces a constant, the eigenvalue, times the original object. The primed basis vector is called an *eigenvector*.

Now, we introduce the special unit tensor $\overline{\overline{1}}$, which is defined as

$$\overline{\overline{1}} = \delta_{ij}\hat{e}_i\hat{e}_j \tag{4.58}$$

so that

$$\overline{\overline{1}} \cdot \overline{V} = \overline{V}. \tag{4.59}$$

Represented as a matrix, $\overline{\overline{1}}$ is simply

$$\overline{\overline{1}} \rightarrow [1] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.60}$$

Using the $\overline{\overline{1}}$ tensor, Equation 4.57 can be rearranged to read

$$\left(\overline{\overline{\sigma}} - \lambda_1\overline{\overline{1}}\right) \cdot \hat{e}_1' = 0. \tag{4.61}$$

Expressing $\overline{\overline{\sigma}}$ in the unprimed system, Equation 4.61 can be written in subscript notation as

$$(\sigma_{ij} - \lambda_1\delta_{ij})\hat{e}_i\hat{e}_j \cdot \hat{e}_1' = 0. \tag{4.62}$$

Equation 4.29 and some rearrangement yields

$$\hat{e}_i(\sigma_{ij} - \lambda_1\delta_{ij})a_{1j} = 0, \tag{4.63}$$

where the a_{1j} are three of the unknown elements of the transformation matrix relating the original coordinate systems to the one in which $\overline{\overline{\sigma}}$ is diagonal.

The LHS of Equation 4.63 is a vector, and for it to be zero, each of its components must be zero. Each component involves a sum over the index j. Equation 4.63 therefore becomes three equations which can be written in matrix array notation as

$$\begin{bmatrix} \sigma_{11} - \lambda_1 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} - \lambda_1 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} - \lambda_1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{4.64}$$

In order for a set of linear, homogeneous equations such as those in Equation 4.64 to have a solution, the determinant of the coefficients must be zero:

$$
\begin{vmatrix}
\sigma_{11} - \lambda_1 & \sigma_{12} & \sigma_{13} \\
\sigma_{21} & \sigma_{22} - \lambda_1 & \sigma_{23} \\
\sigma_{31} & \sigma_{32} & \sigma_{33} - \lambda_1
\end{vmatrix}_{det} = 0. \tag{4.65}
$$

This results in a third-order equation for λ_1 which will generate three eigenvalues. Select one of these values, it does not matter which one because the other two will be used later, and call it λ_1. Inserting this value into Equation 4.64 allows a solution for a_{11}, a_{12}, and a_{13} to within an arbitrary constant. These are three of the elements of the transformation matrix between the primed and unprimed systems that we seek. These three elements also allow the determination of the $\hat{e}_1'$ basis vector to within an arbitrary constant:

$$
\hat{e}_1' = a_{1j}\hat{e}_j. \tag{4.66}
$$

Requiring $\hat{e}_1'$ to be a unit vector determines the arbitrary constant associated with a_{11}, a_{12} and a_{13}:

$$
(a_{11})^2 + (a_{12})^2 + (a_{13})^2 = 1. \tag{4.67}
$$

Except for an overall arbitrary sign and the degenerate situation discussed below, we have now uniquely determined $\hat{e}_1'$.

The other primed basis vectors and elements of the transformation matrix are obtained in a similar way. The second primed basis vector is determined by forming the dot product in Equation 4.56 using $\hat{e}_2'$. A matrix equation equivalent to 4.64 is written with λ_2, a_{21}, a_{22}, and a_{23}. The resulting determinant equation for λ_2 is identical to the one for λ_1, in Equation 4.65. One of the two remaining eigenvalues of this equation is selected for λ_2 and used to determine a_{21}, a_{22}, a_{23}, and $\hat{e}_2'$. In a similar way, the last eigenvalue of Equation 4.65 is used for λ_3 to determine a_{31}, a_{32}, a_{33}, and $\hat{e}_3'$.

The primed coordinate system, in which $\overline{\overline{\sigma}}$ is diagonal, is defined by the basis vectors $\hat{e}_1'$, $\hat{e}_2'$, and $\hat{e}_3'$. The elements of $\overline{\overline{\sigma}}$ in this primed system are just the eigenvalues determined from Equation 4.65,

$$
[\sigma'] = \begin{bmatrix}
\lambda_1 & 0 & 0 \\
0 & \lambda_2 & 0 \\
0 & 0 & \lambda_3
\end{bmatrix}. \tag{4.68}
$$

The matrices of interest in physics and engineering are typically *Hermitian*. If we allow the possibility for complex matrix elements, a matrix is Hermitian if it is equal to its complex conjugate transpose. That is, $\underline{\sigma}_{ij} = \underline{\sigma}_{ji}^*$. There are two important properties of Hermitian matrices. First, their eigenvalues are always pure real numbers. Second, their eigenvectors are always orthogonal. Proofs of these statements are left as exercises at the end of this chapter.

The only complication that can arise in the above diagonalization process is a degenerate situation that occurs when two or more of the eigenvalues are identical. Consider the case when $\lambda_1 \neq \lambda_2 = \lambda_3$. The unique eigenvalue λ_1 determines a_{11},

a_{12}, a_{13} and the eigenvector $\hat{e}_1'$, just as before. The degenerate eigenvalues, however, will not uniquely specify their eigenvectors. These eigenvectors may be chosen an infinite number of ways. An example with this type of degeneracy is discussed in one of the examples that follows.

Example 4.1   As an example of the diagonalization process, consider the conductivity tensor expressed in Cartesian coordinates:

$$\overline{\overline{\sigma}} = \sigma_{ij}\hat{e}_i\hat{e}_j. \tag{4.69}$$

Let this tensor have the matrix representation (ignoring units)

$$[\sigma] = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 10 & 1 \\ 0 & 1 & 10 \end{bmatrix}. \tag{4.70}$$

This matrix is Hermitian, so we can expect to find pure real eigenvalues and orthogonal eigenvectors. The eigenvalues for the diagonalization are generated from the determinant equation

$$\begin{vmatrix} 10 - \lambda & 0 & 0 \\ 0 & 10 - \lambda & 1 \\ 0 & 1 & 10 - \lambda \end{vmatrix}_{det} = 0. \tag{4.71}$$

Expansion of the determinant gives the third-order polynomial equation

$$(10 - \lambda)\left[(10 - \lambda)^2 - 1\right] = 0, \tag{4.72}$$

which has three distinct roots: $\lambda_1 = 9$, $\lambda_2 = 11$, and $\lambda_3 = 10$.

The elements of a_{1j} are determined by inserting the value of λ_1 into Equation 4.64 to obtain

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{4.73}$$

This equation requires $a_{12} = -a_{13}$ and $a_{11} = 0$. The normalization condition imposes the additional constraint that $(a_{12})^2 + (a_{13})^2 = 1$ and results in

$$\begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}. \tag{4.74}$$

The first eigenvector associated with the primed system becomes

$$\hat{e}_1' = (1/\sqrt{2})\hat{e}_2 - (1/\sqrt{2})\hat{e}_3. \tag{4.75}$$

The other components of $[a]$ can be determined by performing similar calculations to give λ_2 and λ_3. The complete transformation matrix is

$$[a] = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 & -1 \\ 0 & 1 & 1 \\ \sqrt{2} & 0 & 0 \end{bmatrix}. \tag{4.76}$$

The other two primed eigenvectors are

$$\hat{e}_2' = (1/\sqrt{2})\hat{e}_2 + (1/\sqrt{2})\hat{e}_3 \tag{4.77}$$

and

$$\hat{e}_3' = \hat{e}_1. \tag{4.78}$$

It should be noted that there is an ordering ambiguity associated with the eigenvalues, as well as a sign ambiguity associated with each eigenvector. These ambiguities allow us to always set up the primed system to be right-handed. The ordering and sign choices made in this example give the primed basis vectors shown in Figure 4.3.

The conductivity tensor elements expressed in the new, diagonal system are

$$[\sigma'] = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 11 & 0 \\ 0 & 0 & 10 \end{bmatrix}. \tag{4.79}$$

Example 4.2 This example demonstrates the diagonalization process when two of the eigenvalues are degenerate. Again consider a conductivity tensor with Cartesian elements (ignoring units)

$$[\sigma] = \begin{bmatrix} 11 & -1 & 0 \\ -1 & 11 & 0 \\ 0 & 0 & 10 \end{bmatrix}. \tag{4.80}$$
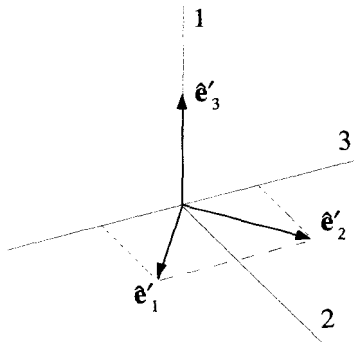

Figure 4.3 The Primed System Basis Vectors

This is a Hermitian matrix and so we expect pure real eigenvalues and orthogonal eigenvectors. The determinant condition is

$$\begin{vmatrix} 11 - \lambda & -1 & 0 \\ -1 & 11 - \lambda & 0 \\ 0 & 0 & 10 - \lambda \end{vmatrix}_{det} = 0, \tag{4.81}$$

which leads to the third-order polynomial equation

$$(10 - \lambda) \left[(11 - \lambda)^2 - 1 \right] . \tag{4.82}$$

This third-order equation has three roots, but only two are distinct, $\lambda_1 = 12$ and $\lambda_2 = \lambda_3 = 10$. The λ_1 root can be treated as before. When substituted into Equation 4.64, the matrix relation becomes

$$\begin{bmatrix} -1 & -1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} . \tag{4.83}$$

This, when coupled with the normalization condition, gives

$$\begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} . \tag{4.84}$$

These elements of the transformation matrix define the first eigenvector

$$\hat{e}_1' = (1/\sqrt{2})\hat{e}_1 - (1/\sqrt{2})\hat{e}_2. \tag{4.85}$$

Now consider the degenerate eigenvalue. When $\lambda_2 = 10$ is substituted into Equation 4.64 we obtain

$$\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_{21} \\ a_{22} \\ a_{23} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} . \tag{4.86}$$

Substitution of λ_3 gives almost the same equation:

$$\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_{31} \\ a_{32} \\ a_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} . \tag{4.87}$$

Equation 4.86 requires $a_{21} = a_{22}$, but gives no constraint on a_{23}. The normalization condition forces the condition $a_{21}^2 + a_{22}^2 + a_{23}^2 = 1$. These conditions can be satisfied by many different eigenvectors. Since a_{23} is arbitrary, set it equal to zero. Now if the second eigenvector is to be orthogonal to $\hat{e}_1'$, we have

$$\begin{bmatrix} a_{21} \\ a_{22} \\ a_{23} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} . \tag{4.88}$$

This gives for the second eigenvector

$$\hat{e}_2' = (1/\sqrt{2})\hat{e}_1 + (1/\sqrt{2})\hat{e}_2. \tag{4.89}$$

The eigenvector associated with λ_3 is given by Equation 4.87 and has all the same constraints as the λ_2 eigenvector, namely, $a_{31} = a_{32}$ and a_{33} is arbitrary. If we want the eigenvectors to be orthogonal, however, $\hat{e}_3'$ must be perpendicular to $\hat{e}_1'$ and $\hat{e}_2'$. The basis vectors $\hat{e}_1'$ and $\hat{e}_2'$ are both in the original 12-plane and so, if $\hat{e}_3'$ is to be perpendicular to these two primed basis vectors, it must be along the 3-direction. This gives

$$\begin{bmatrix} a_{31} \\ a_{32} \\ a_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \tag{4.90}$$

and for the third eigenvector

$$\hat{e}_3' = \hat{e}_3. \tag{4.91}$$

A quick check will show that these three eigenvectors are orthonormal and define a right-handed coordinate system in which the elements of the conductivity tensor are diagonalized.

4.5 TENSOR TRANSFORMATIONS IN CURVILINEAR COORDINATE SYSTEMS

The transformations of the previous sections easily generalize to curvilinear coordinate systems. First consider the intermediate problem of a transformation between a Cartesian and a curvilinear system.

Let the Cartesian system have primed coordinates (x_1', x_2', x_3') and basis vectors $(\hat{e}_1', \hat{e}_2', \hat{e}_3')$, while the unprimed, curvilinear system has coordinates (q_1, q_2, q_3), basis vectors $(\hat{q}_1, \hat{q}_2, \hat{q}_3)$, and scale factors (h_1, h_2, h_3). The set of equations relating the coordinates of the two systems can be written as

$$\begin{aligned} x_1' &= x_1'(q_1, q_2, q_3) \\ x_2' &= x_2'(q_1, q_2, q_3) \\ x_3' &= x_3'(q_1, q_2, q_3). \end{aligned} \tag{4.92}$$

For example, the standard cylindrical system would use the equations

$$\begin{aligned} x_1' &\equiv x' = \rho \cos\theta \\ x_2' &\equiv y' = \rho \sin\theta \\ x_3' &\equiv z' = z. \end{aligned} \tag{4.93}$$

The transformation matrix $[a]$ performs the same function as before. That is, it takes the unprimed, curvilinear components of a vector and generates the primed, Cartesian components:

$$V_i' = a_{ij}V_j. \tag{4.94}$$

Recall from the previous chapter that the displacement vector for the two systems can be written

$$d\bar{\mathbf{r}} = dx_i'\hat{\mathbf{e}}_i' = h_j dq_j \hat{\mathbf{q}}_j. \tag{4.95}$$

The components of the displacement vector in the unprimed curvilinear system are given by the $h_j dq_j$, while its components in the primed Cartesian system are given by the dx_i'. These components must be related by the transformation matrix $[a]$. In subscript notation

$$dx_i' = a_{ij}h_j dq_j. \tag{4.96}$$

The total differential dx_i' can be formed from Equations 4.92 and becomes

$$dx_i' = \frac{\partial x_i'(q_1, q_2, q_3)}{\partial q_j} dq_j. \tag{4.97}$$

Equation 4.97 can be placed in the form of Equation 4.96 by multiplying the RHS of 4.97 by h_j/h_j:

$$dx_i' = \frac{\partial x_i'(q_1, q_2, q_3)}{\partial q_j} \frac{h_j}{h_j} dq_j = \frac{\partial x_i'(q_1, q_2, q_3)}{h_j \partial q_j} h_j dq_j. \tag{4.98}$$

Comparing Equations 4.98 and 4.96 gives

$$a_{ij} = \frac{\partial x_i'(q_1, q_2, q_3)}{h_j \partial q_j} \qquad \text{[Curvilinear} \rightarrow \text{Cartesian]}. \tag{4.99}$$

The further generalization for the transformation between two curvilinear systems follows in a straightforward way. The elements for the transformation matrix $[a]$ in this case become

$$a_{ij} = \frac{h_i' \partial q_i'(q_1, q_2, q_3)}{h_j \partial q_j} \qquad \text{[Curvilinear} \rightarrow \text{Curvilinear]}. \tag{4.100}$$

Note there is no sum over i or j on the RHS of Equation 4.100 because both these subscripts appear on the LHS of the expression.

   Equation 4.100 is the most general form for the elements of the transformation matrix between two curvilinear coordinate systems. It simplifies to Equation 4.99 if the primed system is Cartesian since the $h_i' \rightarrow 1$. It further degenerates to Equation 4.35 if both the primed and unprimed systems are Cartesian since the h_i' and the $h_j \rightarrow 1$.

As before, the transformation matrix can also be determined from the basis vectors of the two coordinate systems. For the general curvilinear case, the elements of $[a]$ are

$$a_{ij} = (\hat{\mathbf{q}}'_i \cdot \hat{\mathbf{q}}_j). \tag{4.101}$$

The above manipulations are fast and easy using subscript notation. It might be a useful exercise to go through the same steps using just matrices to convince yourself the subscript notation is more efficient.

4.6   PSEUDO-OBJECTS

If we consider only transformations that involve rigid rotations or translations, there is no way to change a right-handed system into a left-handed system, or vice-versa. To change handedness requires a reflection. Transformations that involve reflections require the introduction of the so-called "pseudo"-objects. Pseudoscalars, pseudovectors, and pseudotensors are very similar to their "regular" counterparts, except for their behavior when reflected. An easy way to demonstrate the distinction is by closely examining the cross product of two regular vectors in right-handed and left-handed systems. Another way that emphasizes the reflection properties of the transformation is the mirror test, which is presented in Appendix E.

4.6.1   Pseudovectors

Consider the right-handed Cartesian coordinate system shown in Figure 4.4. Picture two regular vectors in this system, oriented along the first two basis vectors:

$$\overline{\mathbf{A}} = A_o \hat{\mathbf{e}}_1 \tag{4.102}$$

$$\overline{\mathbf{B}} = B_o \hat{\mathbf{e}}_2. \tag{4.103}$$

By "regular," we mean that the components of these vectors obey Equation 4.25 when we transform the coordinates.

The cross product between $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ can be formed using the determinant,

$$\overline{\mathbf{A}} \times \overline{\mathbf{B}} = \begin{vmatrix} \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_3 \\ A_o & 0 & 0 \\ 0 & B_o & 0 \end{vmatrix}_{det} = A_o B_o \hat{\mathbf{e}}_3, \tag{4.104}$$

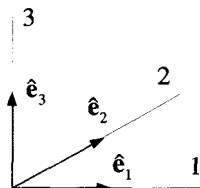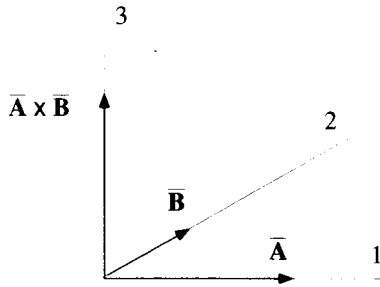**Figure 4.4**   The Right-Handed System

3

$\overline{A} \times \overline{B}$

2

$\overline{B}$

$\overline{A}$ 1

Figure 4.5   Vectors in the Right-Handed System

or, equivalently, using the Levi-Civita symbol:

$$\overline{A} \times \overline{B} = A_i B_j \hat{e}_k \epsilon_{ijk} = A_o B_o \hat{e}_3. \qquad (4.105)$$

The resulting vector is shown in Figure 4.5. Notice how the direction of $\overline{A} \times \overline{B}$ is given by the standard right-hand rule. If you point the fingers of your hand in the direction of $\overline{A}$ and then curl them to point along $\overline{B}$, your thumb will point in the direction of the cross product. Keep in mind that the cross product is not commutative. If the order of the operation is reversed, that is if you form $\overline{B} \times \overline{A}$, the result points in exactly the opposite direction.

Now consider the left-handed system shown in Figure 4.6, with coordinates and basis vectors marked with primes to distinguish them from the coordinates and basis vectors of the right-handed system. This system results from a simple inversion of the 1-axis of the unprimed system. It can also be looked at as a reflection of the right-handed system about its $x_2 x_3$-plane. The equations relating the primed and unprimed coordinates are

$$x_1' = -x_1$$
$$x_2' = +x_2 \qquad (4.106)$$
$$x_3' = +x_3$$

3'

$\hat{e}_3'$ $\hat{e}_2'$ 2'
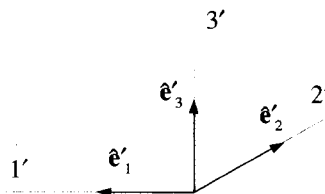
1' $\hat{e}_1'$

Figure 4.6   The Left-Handed System

so that the transformation matrix becomes

$$[a] = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.107}$$

The regular vectors $\overline{A}$ and $\overline{B}$ in the primed coordinate system are simply

$$\overline{A} = -A_o\, \hat{e}_1' \tag{4.108}$$

$$\overline{B} = B_o\, \hat{e}_2'. \tag{4.109}$$

We just wrote these results down because they were obvious. Remember, formally, they can be obtained by applying $[a]$ to the components of the unprimed vectors. The matrix multiplication gives

$$\begin{bmatrix} A_1' \\ A_2' \\ A_3' \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A_o \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -A_0 \\ 0 \\ 0 \end{bmatrix} \tag{4.110}$$

and

$$\begin{bmatrix} B_1' \\ B_2' \\ B_3' \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ B_o \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ B_o \\ 0 \end{bmatrix}. \tag{4.111}$$

It is important to remember that the vectors are the same physical objects in both coordinate systems. They are just being expressed in terms of different components and different basis vectors.

Now form the cross product of $\overline{A}$ and $\overline{B}$ in the left-handed system. To do this we will use the same determinant relation:

$$\overline{A} \times \overline{B} = \begin{vmatrix} \hat{e}_1' & \hat{e}_2' & \hat{e}_3' \\ -A_o & 0 & 0 \\ 0 & B_o & 0 \end{vmatrix}_{det} = -A_o\, B_o\, \hat{e}_3', \tag{4.112}$$

or in terms of the Levi-Civita symbol:

$$\overline{A} \times \overline{B} = A_i'\, B_j'\, \hat{e}_k'\, \epsilon_{ijk} = A_1'\, B_2'\, \hat{e}_3'\, \epsilon_{123} = -A_o\, B_o\, \hat{e}_3'. \tag{4.113}$$

The vectors $\overline{A}$ and $\overline{B}$ and the cross product $\overline{A} \times \overline{B}$ are shown in Figure 4.7 for the left-handed coordinate system. Notice now, the right-hand rule used earlier no longer works to find the direction of the cross product. If we define the cross product using the determinant in Equation 4.112, then we must use a left-hand rule if we are in a left-handed coordinate system.

There is something peculiar here. Compare Figures 4.7 and 4.5 notice that while $\overline{A}$ and $\overline{B}$ point in the same directions in the two systems, their cross product does not!
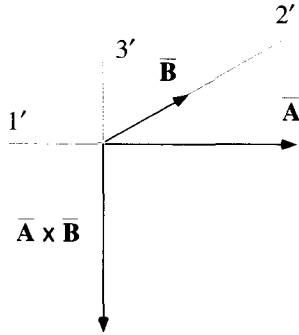
Figure 4.7   Vectors in The Left Handed System

By changing the handedness of the coordinate system, we have managed to change the vector $\overline{A} \times \overline{B}$.

Let's look at this cross product from another point of view. If the unprimed components of the quantity $\overline{A} \times \overline{B}$, given in Equation 4.104, are transformed into the primed system using the $[a]$ matrix, as one would do for regular vector components, we obtain

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ A_o B_o \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ A_o B_o \end{bmatrix}. \tag{4.114}$$

Combining these components with the appropriate basis vectors gives for the cross product vector

$$A_o \, B_o \, \hat{e}'_3. \tag{4.115}$$

This result disagrees with Equation 4.112 by a minus sign. To get around this difficulty, a quantity formed by the cross product of two regular vectors is called a *pseudovector*. Pseudovectors are also commonly called axial vectors, while regular vectors are called polar vectors. If $\overline{V}$ is a regular vector it transforms according to Equation 4.25. However, if $\overline{V}$ is a pseudovector, its components transform according to

$$V'_r = |a|_{\text{det}} \, V_i a_{ri}. \tag{4.116}$$

In this way Equation 4.114 becomes

$$\begin{bmatrix} (\overline{A} \times \overline{B})'_1 \\ (\overline{A} \times \overline{B})'_2 \\ (\overline{A} \times \overline{B})'_3 \end{bmatrix} = - \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ A_o B_o \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -A_o B_o \end{bmatrix} \tag{4.117}$$

giving

$$\overline{A} \times \overline{B} = -A_o \, B_o \, \hat{e}'_3, \tag{4.118}$$

in agreement with Equations 4.112 and 4.113.

To summarize, if $\overline{\mathbf{V}}$ is a regular vector its components transform as

$$V_r' = V_i \, a_{ri}. \tag{4.119}$$

If instead it is a pseudovector, it components transform as

$$V_r' = |a|_{\text{det}} \, V_i a_{ri}. \tag{4.120}$$

If the handedness of two orthonormal coordinate systems is the same, a transformation between them will have $|a|_{\text{det}} = 1$ and vectors and pseudovectors will both transform normally. If the systems have opposite handedness, $|a|_{\text{det}} = -1$ and vectors will transform normally but pseudovectors will flip direction. A vector generated by a cross product of two regular vectors is actually a pseudovector.

It is tempting to think that all this balderdash is somehow a subtle sign error embedded in the definition of the cross product. In some cases, this is correct. For example, when we define the direction of the magnetic field vector, which turns out to be a pseudovector, we have implicitly made an arbitrary choice of handedness that must be treated consistently. Another example is the angular momentum vector, which is defined using a cross product. While you could argue that the "pseudoness" of these two examples is just a problem with their definition, there are cases where you cannot simply explain this property away. It is possible to design situations where an experiment and its mirror image do not produce results which are simply the mirror images of each other. In fact, the Nobel Prize was won by Lee and Yang for analyzing these counterintuitive *violations of parity conservation*. The classic experiment was first performed by Wu, who showed this effect with the emission of beta particles from Cobalt-60, under the influence of the weak interaction.

4.6.2 Pseudoscalars

The ideas that led us to the concept of pseudovectors apply to scalars as well. A proper scalar is invariant to any change of the coordinate system. In contrast, a pseudoscalar changes sign if the handedness of the coordinate system changes. A pseudoscalar involved in a transformation, governed by the transformation matrix $[a]$, will obey

$$S' = |a|_{\text{det}} \, S. \tag{4.121}$$

A good example of a pseudoscalar derives from the behavior of the cross product operation. The volume of a three-dimensional parallelogram, shown in Figure 4.8, can be written as

$$\text{Volume} = (\overline{\mathbf{A}} \times \overline{\mathbf{B}}) \cdot \overline{\mathbf{C}}. \tag{4.122}$$

In a right-handed system, the vector formed by $\overline{\mathbf{A}} \times \overline{\mathbf{B}}$ will point in the upward direction. So in a right-handed system,

$$(\overline{\mathbf{A}} \times \overline{\mathbf{B}}) \cdot \overline{\mathbf{C}} > 0. \tag{4.123}$$

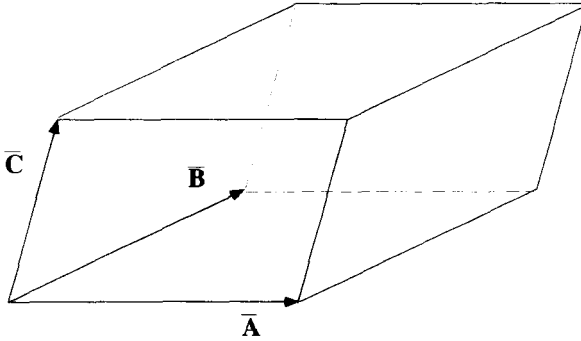**Figure 4.8** The Parallelogram

In a left-handed system, $\overline{\mathbf{A}} \times \overline{\mathbf{B}}$ points downward, and consequently

$$(\overline{\mathbf{A}} \times \overline{\mathbf{B}}) \cdot \overline{\mathbf{C}} < 0. \tag{4.124}$$

Interpreted in this way, the volume of a parallelogram is a pseudoscalar.

4.6.3 Pseudotensors

Pseudotensors are defined just as you would expect. Upon transformation, the components of a pseudotensor obey

$$T'_{rst...} = |a|_{\mathrm{det}}\, T_{ijk...}\, a_{ri}\, a_{sj}\, a_{tk} \cdots, \tag{4.125}$$

which is exactly the same as a regular tensor, except for the $|a|_{\mathrm{det}}$ term.

Again we turn to the cross product to find a good example. Consider two coordinate systems. One, the unprimed system, is a right-handed system, and the other, with primed coordinates, is left-handed. Using the Levi-Civita symbol in both coordinate systems to generate the cross product of $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ gives the relation

$$A_i B_j \hat{e}_k \epsilon_{ijk} = -A'_r B'_s \hat{e}'_t \epsilon'_{rst}. \tag{4.126}$$

The minus sign occurs because, as we showed earlier, the physical direction of the cross product is different in the two coordinate systems. Now the transformation properties of regular vectors can be used to find the relationship between ϵ_{ijk} and ϵ'_{rst}. Because $\overline{\mathbf{A}}$, $\overline{\mathbf{B}}$, and the basis vectors are all regular vectors, they transform according to Equation 4.25. Writing the primed components of these vectors in terms of the unprimed, Equation 4.126 becomes

$$A_i B_j \hat{e}_k \epsilon_{ijk} = -A_i B_j \hat{e}_k a_{ri} a_{sj} a_{tk} \epsilon'_{rst}. \tag{4.127}$$

This expression is true for arbitrary $\overline{A}$ and $\overline{B}$, so we obtain the result

$$\epsilon_{ijk} = -a_{ri}a_{sj}a_{tk}\epsilon'_{rst}. \qquad (4.128)$$

Keep in mind this applies only when the two systems have opposite handedness. If both systems have the same handedness, the minus sign disappears. Thus for the general case of arbitrary transformation between two orthonormal systems, the Levi-Civita symbol components obey

$$\epsilon_{ijk} = |a|_{det}a_{ri}a_{sj}a_{tk}\epsilon'_{rst}. \qquad (4.129)$$

Consequently, the Levi-Civita symbol is a pseudotensor.

EXERCISES FOR CHAPTER 4

1. Determine the transformation matrix $[a]$ that corresponds to a rotation of a two-dimensional Cartesian system by $30°$. Obtain the inverse of this matrix $[a]^{-1}$ and evaluate the following operations:
   (a) $a_{ij}^{-1}a_{jm}$.
   (b) $a_{ij}^{-1}a_{mj}$.
   (c) $a_{ji}^{-1}a_{mj}$.

2. Consider the transformation from a standard two-dimensional Cartesian system (x_1, x_2) to a primed system (x_1', x_2') that results from a reflection about the x_1-axis, as shown below:
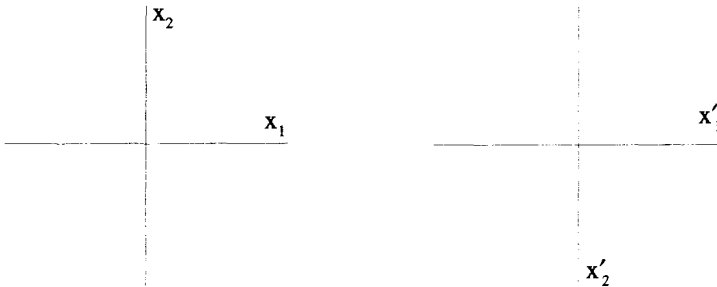
   (a) Express the coordinates of a point in the primed system in terms of the coordinates of the same point in the unprimed system.
   (b) Determine the elements of the transformation matrix $[a]$ that takes vector components from the unprimed system to components in the primed system.
   (c) Determine $[a]^{-1}$ in three ways:
      i. By inverting the $[a]$ matrix found in part (b).
      ii. By inverting the coordinate equations of part (a).
      iii. By simply switching the primed and unprimed labels on the coordinates.

3. Determine the elements of the transformation matrix $[a]$ that generates the Cartesian components of a vector from its components in the toroidal system defined in Exercise 10 of Chapter 3.

4. The coordinates of a hyperbolic system (u, v, z) are related to a set of Cartesian coordinates (x, y, z) by the equations
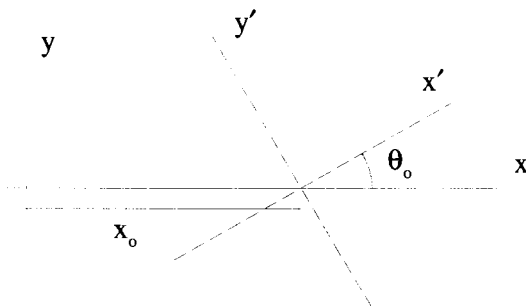
$$u = x^2 - y^2$$

$$v = 2xy$$

$$z = z.$$

Determine the elements of the transformation matrix $[a]$ that takes the Cartesian components of a vector to the hyperbolic components. Using this transformation matrix and the position vector expressed in the Cartesian system, express the position vector in the hyperbolic system.

5. Consider two curvilinear systems, an unprimed cylindrical system and a primed spherical system.
   (a) What are the equations relating the cylindrical coordinates to the spherical coordinates?
   (b) Find the elements of the transformation matrix $[a]$ that generates the primed vector components from unprimed components.

6. Consider a two-dimensional, primed Cartesian system that is shifted an amount x_o and rotated an amount θ_o with respect to a two-dimensional unprimed Cartesian system, as shown in the figure below.

   (a) Identify the equations that generate the xy-coordinates from the $x'y'$-coordinates.
   (b) Determine the elements of the transformation matrix $[a]$ that generate the primed components of a vector from the unprimed components.

7. Consider a two-dimensional Cartesian xy-system and a shifted polar $\rho'\theta'$-system, as shown in the figure below. The origin of the shifted polar system is located at $x = r_o, y = 0$.
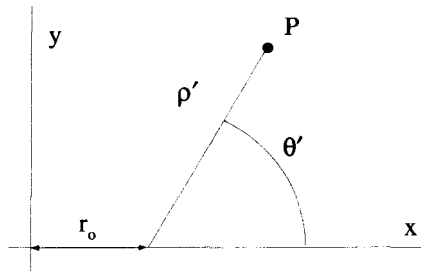
(a) In a sketch, pick a point P well off the x-axis and draw the Cartesian and shifted polar basis vectors.

(b) Express the Cartesian coordinates in terms of the shifted polar coordinates. Invert these equations and express the shifted polar coordinates in terms of the Cartesian coordinates.

(c) Find the elements of the transformation matrix $[a]$ that relates the shifted polar basis vectors to the Cartesian basis vectors.

(d) Express the displacement vector $d\bar{r}$ first in the Cartesian system and then in the shifted polar system.

8. A two-dimensional (u, v) elliptical coordinate system can be related to an (x, y) Cartesian system by the coordinate equations

$$x = \cosh u \cos v$$
$$y = \sinh u \sin v.$$

Determine the elements of the transformation matrix $[a]$ that converts vector components in the Cartesian system to vector components in this elliptical system. What is $[a]^{-1}$?

9. Consider three coordinate systems: (1) a Cartesian system with coordinates (x_1, x_2, x_3) and basis vectors $(\hat{e}_1, \hat{e}_2, \hat{e}_3)$; (2) a curvilinear system with coordinates (q_1, q_2, q_3) and basis vectors $(\hat{q}_1, \hat{q}_2, \hat{q}_3)$; and (3) a second curvilinear system with coordinates (q_1', q_2', q_3') and basis vectors $(\hat{q}_1', \hat{q}_2', \hat{q}_3')$. Let the relationships between the Cartesian coordinates and these curvilinear coordinates be given by

$$\begin{aligned}
x &= x(q_1, q_2, q_3) & x &= x'(q_1', q_2', q_3') \\
y &= y(q_1, q_2, q_3) & y &= y'(q_1', q_2', q_3') \\
z &= z(q_1, q_2, q_3) & z &= z'(q_1', q_2', q_3').
\end{aligned} \qquad (4.130)$$

(a) Find the general expressions for the elements of the transformation matrix $[a]$ that takes vector components from one curvilinear system to the other.

(b) Take one of the curvilinear systems to be the cylindrical system, and the other to be the elliptical system of Exercise 8 and specifically determine the elements of $[a]$ and its inverse.

10. Consider a two-dimensional Cartesian system with coordinates $(x \rightarrow x_1, y \rightarrow x_2)$ and basis vectors $(\hat{e}_x \rightarrow \hat{e}_1, \hat{e}_y \rightarrow \hat{e}_2)$ and a polar system with coordinates $(\rho \rightarrow q_1, \phi \rightarrow q_2)$ and basis vectors $(\hat{e}_\rho \rightarrow \hat{q}_1, \hat{e}_\phi \rightarrow \hat{q}_2)$.

   (a) Express the Cartesian coordinates (x_1, x_2) in terms of the polar coordinates (q_1, q_2).

   (b) Express the position vector $\bar{r}$, first using the Cartesian basis vectors and then using the polar basis vectors.

   (c) Determine the elements of the transformation matrix $[a]$ that takes vector components from the polar system to vector components in the Cartesian system. Show that this transformation matrix works for the expressions for the position vector of part (b) above.

   (d) What are the elements of $[a]^{-1}$, the matrix that generates polar components from the Cartesian components?

   (e) Determine the elements of the displacement vector $d\bar{r}$ in the Cartesian and polar systems.

   (f) A second-rank tensor expressed in the Cartesian system takes the following form

$$\overline{\overline{T}} = \sum_{i=1}^{2} \sum_{j=1}^{2} 1 \hat{e}_i \hat{e}_j.$$

   What are its elements in the polar system?

11. Consider the transformation from a two-dimensional Cartesian 12-system to a Cartesian $1'2'$-system, which includes both an inversion and a rotation as shown in the drawing below.
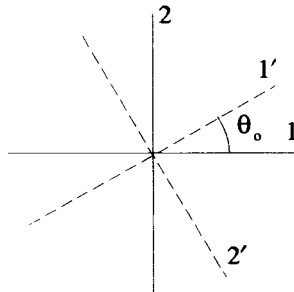
   (a) Express the vector $\overline{V} = 3\hat{e}_1 + 2\hat{e}_2$ in the $1'2'$-system.

   (b) Express $\overline{\overline{T}} = ij\hat{e}_i\hat{e}_j$ in the $1'2'$-system. Do not forget to sum over the repeated i and j subscripts.

12. Consider a two-dimensional conductivity tensor $\overline{\overline{\sigma}}$ whose elements expressed in an (x, y) Cartesian system have the matrix representation

$$\overline{\overline{\sigma}} \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

The current density $\overline{J}$ is related to the electric field $\overline{E}$ through Ohm's law:

$$\overline{J} = \overline{\overline{\sigma}} \cdot \overline{E}.$$

(a) If the electric field is given by $\overline{E} = E_o \hat{e}_x + E_o \hat{e}_y$, express the current density in this Cartesian system.

(b) Now transform the components of $\overline{\overline{\sigma}}$ to a new orthogonal (x', y') system, where the basis vector $\hat{e}'_x$ is parallel to the $\overline{E}$ given in part (a) above. The electric field in this new system is given as $\overline{E} = \sqrt{2}E_o \hat{e}'_x$. What are the elements of the transformation matrix that accomplishes this? What are the components of $\overline{\overline{\sigma}}$ in this new system?

(c) With $\overline{\overline{\sigma}}$ and $\overline{E}$ expressed in the primed system, form the product $\overline{\overline{\sigma}} \cdot \overline{E}$ in that system and show that it results in the same vector obtained in part (a) above.

13. Show, using matrix array and subscript/summation notation, that in general

$$\overline{\overline{\sigma}} \cdot \overline{E} \neq \overline{E} \cdot \overline{\overline{\sigma}}.$$

14. Consider the inner product of two second-rank tensors, each expressed in the same orthonormal system:

$$\overline{\overline{A}} \cdot \overline{\overline{B}}.$$

In subscript/summation notation this product involves the term

$$\hat{e}_i \hat{e}_j \cdot \hat{e}_m \hat{e}_n.$$

(a) Develop an expression for $\overline{\overline{A}} \cdot \overline{\overline{B}}$ using the Kronecker-δ and show that the result is also a second-rank tensor.

(b) Show that $\overline{\overline{A}} \cdot \overline{\overline{B}}$ does not equal $\overline{\overline{B}} \cdot \overline{\overline{A}}$.

15. Let $\overline{\overline{T}}$ be a second-rank tensor and $\overline{V}$ be a vector. Expressed in a Cartesian system,

$$\overline{\overline{T}} = T_{ij} \hat{e}_i \hat{e}_j$$
$$\overline{V} = V_i \hat{e}_i.$$

(a) Express the following using subscript/summation notation and, where necessary, the Levi-Civita symbol

    i. $\overline{V} \cdot \overline{\overline{T}}$.

    ii. $\overline{\overline{T}} \cdot \overline{V}$.

    iii. $(\overline{V} \cdot \overline{\overline{T}}) \times \overline{V}$.

    iv. $(\overline{V} \cdot \overline{\overline{T}}) \times (\overline{\overline{T}} \cdot \overline{V})$.

    v. $(\overline{V} \cdot \overline{\overline{T}}) \cdot (\overline{\overline{T}} \cdot \overline{V})$.

(b) What values for the elements of $\overline{\overline{T}}$, other than $T_{ij} = 0$, will result in $\overline{V} \times (\overline{V} \cdot \overline{\overline{T}}) = 0$ for any $\overline{V}$?

(c) What values for the elements of $\overline{\overline{T}}$, other than $T_{ij} = 0$, will result in $\overline{V} \cdot (\overline{V} \cdot \overline{\overline{T}}) = 0$ for any $\overline{V}$?

16. Use subscript/summation notation to show that the expressions

$$\left(\overline{\overline{T}} \cdot \overline{C}\right) \times \overline{D} \qquad \text{and} \qquad \overline{\overline{T}} \cdot \left(\overline{C} \times \overline{D}\right)$$

are not necessarily equal.

17. An orthogonal (u, v, z) coordinate system is defined by the set of equations relating its coordinates to a standard set of Cartesian coordinates,

$$u = \frac{x}{x^2 + y^2}$$

$$v = \frac{-y}{x^2 + y^2}$$

$$z = z.$$

(a) Determine the elements of the transformation matrix $[a]$.

(b) At the Cartesian point $(1, 1, 1)$ let the vector $\overline{V} = 1\,\hat{e}_x + 2\,\hat{e}_y + 3\,\hat{e}_z$. Express $\overline{V}$ at this point using (u, v, z) system vector components and basis vectors.

(c) At the Cartesian point $(1, 1, 1)$ let the tensor $\overline{\overline{T}} = 1\,\hat{e}_1\hat{e}_3 + 2\,\hat{e}_2\hat{e}_1 + 3\,\hat{e}_2\hat{e}_3 + 4\,\hat{e}_3\hat{e}_2$. Express $\overline{\overline{T}}$ at this point using (u, v, z) system tensor elements and basis vectors.

18. Consider a dumbbell positioned in the xy-plane of a Cartesian system, as shown in the figure below. The moment of inertia tensor for this object expressed in this Cartesian system is

$$\overline{\overline{I}} = I_{ij}\hat{e}_i\hat{e}_j$$

$$\overline{\overline{I}} \rightarrow M_o \begin{bmatrix} 2 & -2 & 0 \\ -2 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

Find the basis vectors of a coordinate system in which the moment of inertia tensor is diagonalized. Draw the dumbbell in that system.
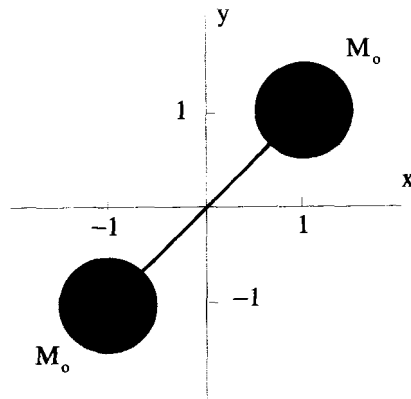
19. Let a tensor $\overset{=}{\mathbf{T}} = T_{ij}\hat{\mathbf{e}}_i\hat{\mathbf{e}}_j$ in a two-dimensional Cartesian system be represented by the matrix array

$$\overset{=}{\mathbf{T}} \to \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}.$$

(a) Find the eigenvalues and eigenvectors of this matrix array.

(b) Plot the eigenvectors and show their relation to the Cartesian basis vectors.

(c) Are the eigenvectors orthogonal? Is the matrix array Hermitian?

(d) Repeat parts (a)–(c) for the tensor

$$\overset{=}{\mathbf{T}} \to \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}.$$

20. Let the components of the conductivity tensor in a Cartesian system be represented by

$$\overset{=}{\boldsymbol{\sigma}} \to [\sigma] = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Identify an electric field vector, by specifying its components, that will result in a current density that is parallel to this electric field.

21. Using subscript/summation notation, show that the eigenvalues of a Hermitian matrix are pure real. Show that the eigenvectors generated by the diagonalization of a Hermitian matrix generate orthogonal eigenvectors.

Start with a second-rank tensor that is assumed to have off-diagonal elements that may be complex,

$$\overset{=}{\mathbf{T}} = \underline{T}_{ij}\hat{\mathbf{e}}_i\hat{\mathbf{e}}_j.$$

Let the eigenvalues of the matrix representation of this tensor be $\underline{\lambda}_i$ and the eigenvectors associated with these eigenvalues be $\hat{e}'_i = \underline{a}_{ij}\hat{e}_j$ so that

$$\underline{T}_{ij}\underline{a}_{nj} = \underline{\lambda}_n\underline{a}_{ni}.$$

Notice that although the n subscript is repeated on the RHS it is not summed over because it appears on the LHS, where it is clearly not summed over. Now form the quantities $\underline{a}^*_{mi}\underline{T}_{ij}\underline{a}_{nj}$ and $\underline{a}^*_{ni}\underline{T}_{ij}\underline{a}_{mj}$ and subtract one from the other. Apply the Hermitian condition and see what it implies.

22. Is the dot product between two pseudovectors a scalar or a pseudoscalar? Give an example from classical mechanics of such an operation.

23. Show that the elements of the Kronecker delta δ_{ij} transform like elements of a tensor, not a pseudotensor.

5

THE DIRAC δ-FUNCTION

The Dirac δ-function is a strange, but useful function which has many applications in science, engineering, and mathematics. The δ-function was proposed in 1930 by Paul Dirac in the development of the mathematical formalism of quantum mechanics. He required a function which was zero everywhere, except at a single point, where it was discontinuous and behaved like an infinitely high, infinitely narrow spike of unit area. Mathematicians were quick to point out that, strictly speaking, there is no function which has these properties. But Dirac supposed there was, and proceeded to use it so successfully that a new branch of mathematics was developed in order to justify its use. This area of mathematics is called the theory of *generalized functions* and develops, in complete detail, the foundation for the Dirac δ-function. This rigorous treatment is necessary to justify the use of these discontinuous functions, but for the physicist the simpler physical interpretations are just as important. We will take both approaches in this chapter.

5.1   EXAMPLES OF SINGULAR FUNCTIONS IN PHYSICS

Physical situations are usually described using equations and operations on continuous functions. Sometimes, however, it is useful to consider discontinuous idealizations, such as the mass density of a point mass, or the force of an infinitely fast mechanical impulse. The functions that describe these ideas are obviously extremely discontinuous, because they and all their derivatives must diverge. For this reason they are often called *singular* functions. The Dirac δ-function was developed to describe functions that involve these types of discontinuities and provide a method for handling them in equations which normally involve only continuous functions.
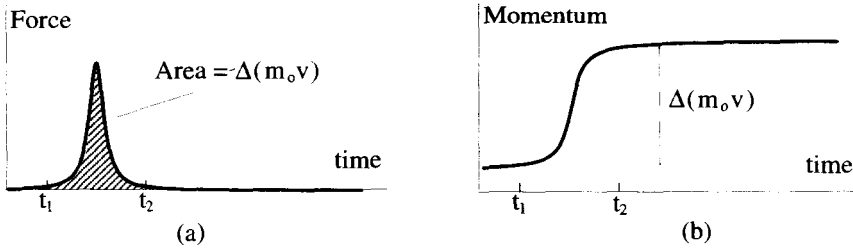
Figure 5.1   Force and Change in Momentum

5.1.1   The Ideal Impulse

Often a student's first encounter with the δ-function is the "ideal" impulse. In mechanics, an impulse is a force which acts on an object over a finite period of time. Consider the realistic force depicted in Figure 5.1(a). It is zero until $t = t_1$, when it increases smoothly from zero to its peak value, and then finally returns back to zero at $t = t_2$. When this force is applied to an object of mass m_o, the momentum in the direction of the applied force changes, as shown in Figure 5.1(b). The momentum remains constant until $t = t_1$, when it begins to change continuously until reaching its final value at $t = t_2$. The net momentum change $\Delta(m_o v)$ is equal to the integrated area of the force curve:

$$\int_{-\infty}^{+\infty} dt\, F(t) = \int_{t_1}^{t_2} dt\, F(t)$$

$$= \int_{t_1}^{t_2} dt\, m_o \frac{dv}{dt} = m_o v \Big|_{t_2} - m_o v \Big|_{t_1} = \Delta(m_o v). \qquad (5.1)$$

An ideal impulse produces all of its momentum change instantaneously, at the single point $t = t_o$, as shown in Figure 5.2(a). Of course this is not very realistic, since it requires an infinite force to change the momentum of a finite mass in zero time. But it is an acceptable thought experiment, because we might be considering the limit in which a physical process occurs faster than any measurement can detect.
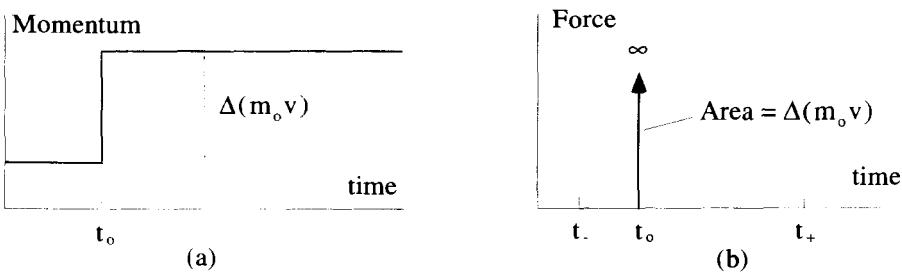

Figure 5.2   An Instantaneous Change in Momentum

The force of an ideal impulse cannot be graphed as a function of time in the normal sense. The force exists for only a single instant, and so is zero everywhere, except at $t = t_o$, when it is infinite. But this is not just any infinity. Since the total momentum change must be $\Delta(m_o v)$, the force must diverge so that its integral area obeys (for all $t_- < t_+$)

$$\int_{t_-}^{t_+} dt\, F(t) = \begin{cases} \Delta(m_o v) & t_- < t_o < t_+ \\ 0 & \text{otherwise} \end{cases} . \qquad (5.2)$$

In other words, any integral which includes the point t_o gives a momentum change of $\Delta(m_o v)$. On the other hand, integrals which exclude t_o must give no momentum change. We graph this symbolically as shown in Figure 5.2(b). A spike of zero width with an arrow indicates the function goes to infinity, while the area of the impulse is usually indicated by a comment on the graph, as shown in the figure, or sometimes by the height of the arrow.

The Dirac δ-function $\delta(t)$ was designed to represent exactly this kind of "pathological" function. $\delta(t)$ is zero everywhere, except at $t = 0$, when it is infinite. Again, this is not just any infinity. It diverges such that any integral area which includes $t = 0$ has the value of 1. That is (for all $t_- < t_+$),

$$\int_{t_-}^{t_+} dt\, \delta(t) = \begin{cases} 1 & t_- < 0 < t_+ \\ 0 & \text{otherwise} \end{cases} . \qquad (5.3)$$

The symbolic plot is shown in Figure 5.3. The ideal impulse, discussed above, can be expressed in terms of a *shifted* Dirac δ-function:

$$F(t) = \Delta(m_o v)\delta(t - t_o). \qquad (5.4)$$

The $t - t_o$ argument simply translates the spike of the δ-function so that it occurs at t_o instead of 0.

5.1.2   Point Masses and Point Charges

Physical equations often involve the mass per unit volume $\rho_m(\bar{r})$ of a region of space. Normally $\rho_m(\bar{r})$ is a continuous function of position, but with δ-functions, it can also represent point masses. A point mass has a finite amount of mass stuffed inside a single point of space, so the density must be infinite at that point and zero everywhere else.
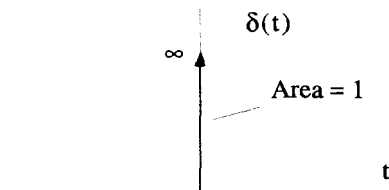

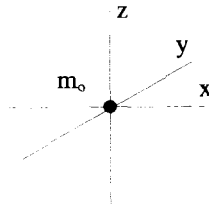**Figure 5.3**   The Dirac δ-function, $\delta(t)$

Figure 5.4   A Single Point Mass at the Origin

Integrating the mass density over a volume V gives the total mass enclosed:

$$\int_V d\tau \, \rho_m(\overline{\mathbf{r}}) = \text{ total mass inside } V. \tag{5.5}$$

Thus, if there is a single point of mass m_o located at the origin, as shown in Figure 5.4, any volume integral which includes the origin must give the total mass as m_o. Integrals which exclude the origin must give zero. In mathematical terms:

$$\int_V d\tau \, \rho_m(\overline{\mathbf{r}}) = \begin{cases} m_o & \text{origin included in } V \\ 0 & \text{origin excluded from } V \end{cases}. \tag{5.6}$$

Using Dirac δ-functions, this mass density function becomes

$$\rho_m(\overline{\mathbf{r}}) = m_o \delta(x) \delta(y) \delta(z). \tag{5.7}$$

Equation 5.6 can easily be checked by expanding the integral as

$$\int_V d\tau \, \rho_m(\overline{\mathbf{r}}) = \int dx \int dy \int dz \, m_o \delta(x) \delta(y) \delta(z). \tag{5.8}$$

Application of Equation 5.3 on the three independent integrals gives m_o only when V includes the origin. If, instead of the origin, the point mass is located at the point (x_o, y_o, z_o), shifted arguments are used in each of the δ-functions:

$$\rho_m(\overline{\mathbf{r}}) = m_o \delta(x - x_o) \delta(y - y_o) \delta(z - z_o). \tag{5.9}$$

Dirac δ-functions can also be used, in a similar way, to represent point charges in electromagnetism.

5.2   TWO DEFINITIONS OF $\delta(t)$

There are two common ways to define the Dirac δ-function. The more rigorous approach, from the theory of generalized functions, defines it by its behavior inside integral operations. In fact, the δ-function is actually never supposed to exist outside an integral. In general, scientists and engineers are a bit more lax and use a second definition. They often define the δ-function as the limit of an infinite sequence of
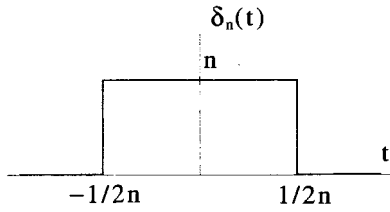
Figure 5.5   The Square Sequence Function

continuous functions. Also, as demonstrated in the two examples of the previous section, they frequently manipulate the δ-function outside of integrals. Usually, there are no problems with this less rigorous approach. However, there are some cases, such as the oscillatory sequence functions described at the end of this section, where the more careful integral approach becomes essential.

5.2.1 $\delta(t)$ as the Limit of a Sequence of Functions

The δ-function can be viewed as the limit of a sequence of functions. In other words,

$$\delta(t) = \lim_{n \to \infty} \delta_n(t), \tag{5.10}$$

where $\delta_n(t)$ is finite for all values of t.

There are many function sequences that approach the Dirac δ-function in this way. The simplest is the square function sequence defined by

$$\delta_n(t) = \begin{cases} n & -1/2n < t < +1/2n \\ 0 & \text{otherwise} \end{cases} \tag{5.11}$$

and shown in Figure 5.5. Clearly, for any value of n

$$\int_{-\infty}^{\infty} dt \, \delta_n(t) = 1, \tag{5.12}$$

and in the limit as $n \to \infty$, $\delta_n(t) = 0$ for all t, except $t = 0$. The first three square sequence functions for $n = 1, 2,$ and 3 are shown in Figure 5.6.
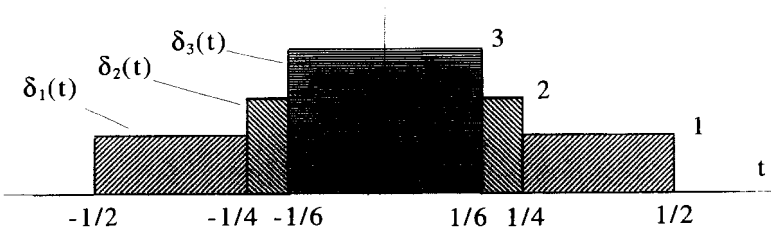

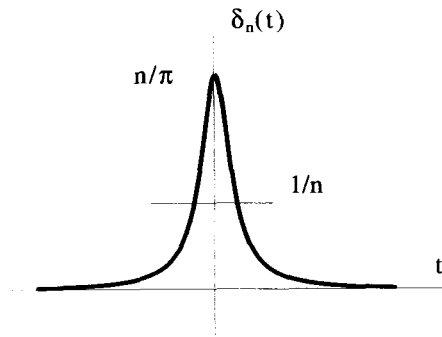**Figure 5.6**   The First Three Square Sequence Functions

Figure 5.7   The Resonance Sequence Function

There are four other common function sequences that approach the Dirac δ-function. Three of them, the resonance, Gaussian, and sinc squared sequences, are shown in Figures 5.7–5.9 and are mathematically described by

$$\text{Resonance:} \quad \delta_n(t) = \frac{n/\pi}{1 + n^2 t^2}$$

$$\text{Gaussian:} \quad \delta_n(t) = \frac{n}{\sqrt{\pi}} e^{-n^2 t^2} \tag{5.13}$$

$$\text{Sinc Squared:} \quad \delta_n(t) = \frac{\sin^2 nt}{n\pi t^2}.$$

Each of these functions has unit area for any value of n, and it is easy to calculate that in the limit as $n \to \infty$, $\delta_n(t) = 0$ for all $t \neq 0$.

There is one other common sequence, the sinc function sequence, which approaches the δ-function in a completely different manner. The discussion of this requires more rigor, and is deferred until the end of this section.
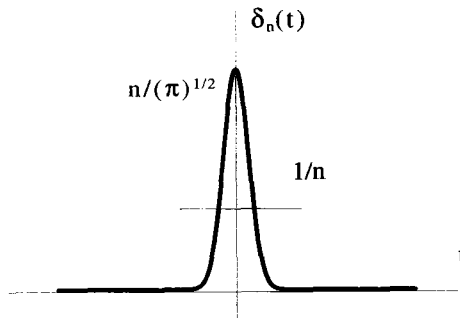

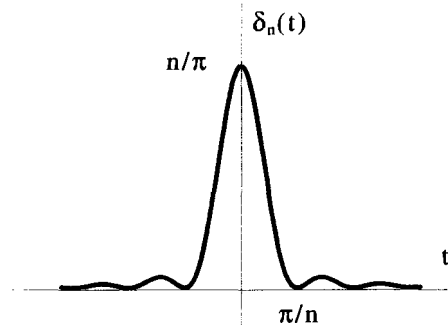**Figure 5.8**   The Gaussian Sequence Function

Figure 5.9   The Sinc Squared Sequence Function

5.2.2   Defining δ(*t*) by Integral Operations

In mathematics, the δ-function is defined by how it behaves inside an integral. Any function which behaves as $\delta(t)$ in the following equation is *by definition* a δ-function,

$$\int_{t_-}^{t_+} dt\; \delta(t - t_o)f(t) = \begin{cases} f(t_o) & t_- < t_o < t_+ \\ 0 & \text{otherwise} \end{cases}, \qquad (5.14)$$

where $t_- < t_+$ and $f(t)$ is any continuous, well-behaved function. This operation is sometimes called a *sifting integral* because it selects the single value $f(t_o)$ out of $f(t)$.

Because this is a definition, it need not be proven, but its consistency with our previous definition and applications of the δ-function must be shown. If $t_- < t_o < t_+$, the range of the integral can be changed to be an infinitesimal region of size 2ϵ centered around t_o, without changing the value of the integral. This is true because $\delta(t - t_o)$ vanishes everywhere except at $t = t_o$. This means

$$\int_{t_-}^{t_+} dt\; \delta(t - t_o)f(t) = \int_{t_o - \epsilon}^{t_o + \epsilon} dt\; \delta(t - t_o)f(t). \qquad (5.15)$$

Because $f(t)$ is a continuous function, over the infinitesimal region it is effectively a constant, with the value $f(t_o)$. Therefore,

$$\int_{t_-}^{t_+} dt\; \delta(t - t_o)f(t) = f(t_o) \int_{t_o - \epsilon}^{t_o + \epsilon} dt\; \delta(t - t_o) = f(t_o). \qquad (5.16)$$

This "proves" the first part of Equation 5.14. The second part, when t_o is not inside the range $t_- < t < t_+$, easily follows because, in this case, $\delta(t - t_o)$ is zero for the entire range of the integrand.

Figure 5.10 shows a representation of the integration in Equation 5.14. The integrand is a product of a shifted δ-function and the continuous function $f(t)$. Because the δ-function is zero everywhere except at $t = t_o$, the integrand goes to a δ-function located at $t = t_o$, with an area scaled by the value of $f(t_o)$.
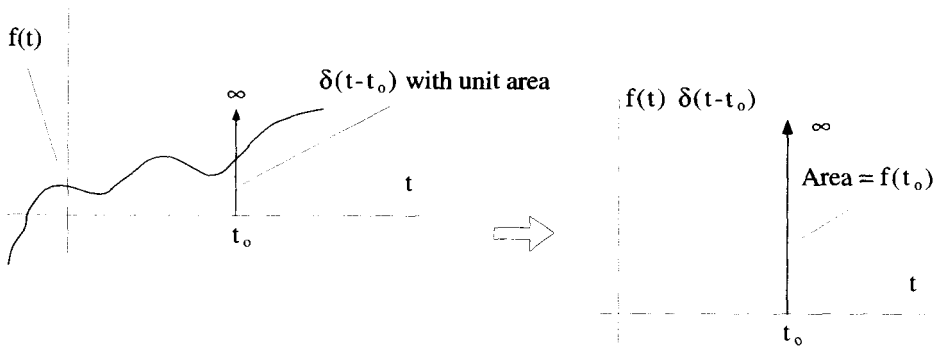
Figure 5.10   A Graphical Interpretation of the Sifting Integral

5.2.3   The Sinc Sequence Function

The sinc function can be used to form another set of sequence functions which, in the limit as $n \to \infty$, approach the behavior of a δ-function. The sinc function sequence is shown in Figure 5.11 and is defined by

$$\text{Sinc:} \quad \delta_n(t) = \frac{\sin nt}{\pi t}. \tag{5.17}$$

Notice, however, that as $n \to \infty$ this $\delta_n(t)$ does not approach zero for all $t \neq 0$, so the sinc sequence does not have the characteristic infinitely narrow, infinitely tall peak that we have come to expect. How, then, can we claim this sequence approaches a δ-function?

The answer is that the sinc function approaches the δ-function behavior from a completely different route, which can only be understood in the context of the integral definition of the δ-function. In the limit as $n \to \infty$, the sinc function oscillates infinitely fast except at $t = 0$. Thus, when the sequence is applied to a continuous function $f(t)$ in a sifting integral, the only contribution which is not canceled out by rapid oscillations comes from the point $t = 0$. The result is, as you will prove in one of the
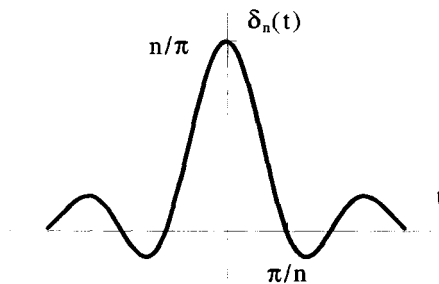

Figure 5.11   The Sinc Sequence Function

exercises at the end of this chapter, the same as you would expect for any δ-function:

$$\int_{t_-}^{t_+} dt \left[\lim_{n\to\infty} \frac{\sin nt}{\pi t}\right] f(t) = \begin{cases} f(0) & t_- < 0 < t_+ \\ 0 & \text{otherwise} \end{cases}. \tag{5.18}$$

This means, in the limit $n \to \infty$, the sinc sequence approaches the δ-function:

$$\lim_{n\to\infty} \frac{\sin nt}{\pi t} = \delta(t). \tag{5.19}$$

5.3 δ-FUNCTIONS WITH COMPLICATED ARGUMENTS

So far, we have only considered the Dirac δ-function with either a single, independent variable as an argument, e.g., $\delta(t)$, or the shifted version $\delta(t - t_o)$. In general, however, the argument could be any function of the independent variable (or variables). It turns out that such a function can always be rewritten as a sum of simpler δ-functions. Three specific examples of how this is accomplished are presented in this section, and then the general case is explored. In each case, the process is the same. The complicated δ-function is inserted into an integral, manipulated, and then rewritten in terms of δ-functions with simpler arguments.

5.3.1 δ(−t)

To determine the properties of $\delta(-t)$, make it operate on any continuous function $f(t)$ inside a sifting integral,

$$\int_{t_-}^{t_+} dt\, \delta(-t) f(t), \tag{5.20}$$

where $t_- < t_+$. The substitution $t' = -t$ transforms this to

$$-\int_{-t_-}^{-t_+} dt'\, \delta(t') f(-t') = \int_{-t_+}^{-t_-} dt'\, \delta(t') f(-t')$$

$$= \begin{cases} f(0) & -t_+ < 0 < -t_- \\ 0 & \text{otherwise} \end{cases}, \tag{5.21}$$

where the last step follows from the integral definition of $\delta(t)$ in Equation 5.14. Therefore, we have

$$\int_{t_-}^{t_+} dt\, \delta(-t) f(t) = \begin{cases} f(0) & t_- < 0 < t_+ \\ 0 & \text{otherwise} \end{cases}. \tag{5.22}$$

But notice this is precisely the result obtained if a $\delta(t)$ were applied to the same function:

$$\int_{t_-}^{t_+} dt\ \delta(t)f(t) = \begin{cases} f(0) & t_- < 0 < t_+ \\ 0 & \text{otherwise} \end{cases}. \tag{5.23}$$

Remember, by definition, anything that behaves like a δ-function inside an integral is a δ-function, so $\delta(-t) = \delta(t)$. This implies the δ-function is an even function.

5.3.2 $\delta(at)$

Now consider another simple variation $\delta(at)$, where a is any positive constant. Again, start with a sifting integral in the form

$$\int_{t_-}^{t_+} dt\ f(t)\delta(at). \tag{5.24}$$

With a variable change of $t' = at$, this becomes

$$\int_{t_-/a}^{t_+/a} dt'\ \frac{f(t'/a)\delta(t')}{a} = \begin{cases} f(0)/a & t_-/a < 0 < t_+/a \\ 0 & \text{otherwise} \end{cases}, \tag{5.25}$$

so that

$$\int_{t_-}^{t_+} dt\ f(t)\delta(at) = \begin{cases} f(0)/a & t_- < 0 < t_+ \\ 0 & \text{otherwise} \end{cases}. \tag{5.26}$$

Notice again, this is just the definition of the function $\delta(t)$ multiplied by the constant $1/a$, so

$$\delta(ax) = \delta(x)/a \qquad a > 0. \tag{5.27}$$

This derivation was made assuming a positive a. The same manipulations can be performed with a negative a, and combined with the previous result, to obtain the more general expression

$$\delta(ax) = \delta(x)/|a|. \tag{5.28}$$

Notice that our first example, $\delta(t) = \delta(-t)$, can be derived from Equation 5.28 by setting $a = -1$.

5.3.3 $\delta(t^2 - a^2)$

As a bit more complicated argument for the Dirac δ-function, consider $\delta(t^2 - a^2)$. The argument of this function goes to zero when $t = +a$ and $t = -a$, which seems to imply two δ-functions. To test this theory, place the function in the sifting integral

$$\int_{t_-}^{t_+} dt\ f(t)\delta(t^2 - a^2). \tag{5.29}$$

There can be contributions to this integral only at the zeros of the argument of the δ-function. Assuming that the integral range includes both these zeros, i.e., $t_- < -a$ and $t_+ > +a$, this integral becomes

$$\int_{-a-\epsilon}^{-a+\epsilon} dt\, f(t)\delta(t^2 - a^2) + \int_{+a-\epsilon}^{+a+\epsilon} dt\, f(t)\delta(t^2 - a^2), \qquad (5.30)$$

which is valid for any value of $0 < \epsilon < a$.

Now $t^2 - a^2 = (t - a)(t + a)$, which near the two zeros can be approximated by

$$t^2 - a^2 = (t - a)(t + a) \approx \begin{cases} (t + a)(-2a) & t \to -a \\ (t - a)(+2a) & t \to +a \end{cases}. \qquad (5.31)$$

Formally, these results are obtained by performing a Taylor series expansion of $t^2 - a^2$ about both $t = -a$ and $t = +a$ and keeping terms up to the first order. The Taylor series expansion of $t^2 - a^2$ around an arbitrary point t_o is, to first order, given by

$$t^2 - a^2 \approx t_o^2 - a^2 + \left.\frac{d(t^2 - a^2)}{dt}\right|_{t=t_o} (t - t_o). \qquad (5.32)$$

In the limit as $\epsilon \to 0$, Integral 5.30 then becomes

$$\int_{-a-\epsilon}^{-a+\epsilon} dt\, f(t)\delta\big(-2a(t + a)\big) + \int_{+a-\epsilon}^{+a+\epsilon} dt\, f(t)\delta\big(2a(t - a)\big). \qquad (5.33)$$

Using the result from the previous section, $\delta(at) = \delta(t)/|a|$, gives

$$\int_{t_-}^{t_+} dt\, f(t)\delta(t^2 - a^2) = \frac{1}{2a}f(-a) + \frac{1}{2a}f(+a). \qquad (5.34)$$

Therefore, $\delta(t^2 - a^2)$ is equivalent to the sum of two δ-functions:

$$\delta(t^2 - a^2) = \frac{1}{2a}\delta(t - a) + \frac{1}{2a}\delta(t + a), \qquad (5.35)$$
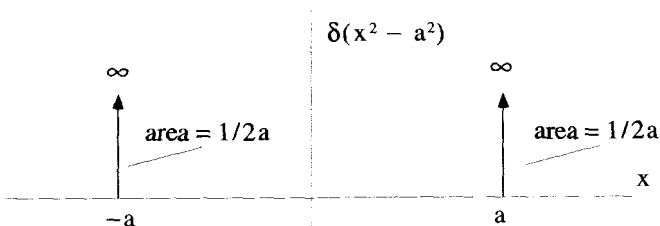
as shown in Figure 5.12.

Figure 5.12   The Plot of $\delta(x^2 - a^2)$

5.3.4 The General Case of $\delta(f(t))$

Using the Taylor series approach, the previous example can be easily generalized to an arbitrary argument inside a δ-function:

$$\delta\big(f(t)\big) = \sum_i \left| \frac{1}{df/dt|_{t=t_i}} \right| \delta(t - t_i), \tag{5.36}$$

where the sum over i is a sum over all the zeros of $f(t)$ and t_i is the value of t where each zero occurs.

5.4 INTEGRALS AND DERIVATIVES OF $\delta(t)$

Integrating the δ-function is straightforward because the δ-function is defined by its behavior inside integrals. But, because the δ-function is extremely discontinuous, you might think that it is impossible to talk about its derivatives. While it is true that the derivatives cannot be treated like those of continuous functions, it is possible to talk about them inside integral operations or as limits of the derivatives of a sequence of functions. These manipulations are sometimes referred to as δ-function calculus.

5.4.1 The Heaviside Unit Step Function

Consider the function $H(x)$, which results from integrating the δ-function:

$$H(x) = \int_{-\infty}^{x} dt \, \delta(t). \tag{5.37}$$

$H(x)$ can be interpreted as the area under $\delta(t)$ in the range from $t = -\infty$ to $t = x$, as shown in Figure 5.13. If $x < 0$, the range of integration does not include $t = 0$, and $H(x) = 0$. As soon as x exceeds zero, the integration range includes $t = 0$, and $H(x) = 1$:

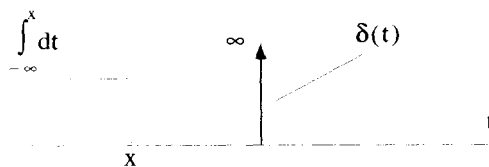$$H(x) = \int_{-\infty}^{x} dt \, \delta(t) = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases}. \tag{5.38}$$

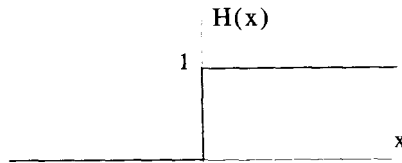**Figure 5.13** The Integration for the Heaviside Step Function

H(x)

Figure 5.14   The Heaviside Function

Strictly, the value of $H(0)$ is not well defined, but because $\delta(t)$ is an even function, many people define $H(0) = 1/2$. A plot of the Heaviside function is shown in Figure 5.14.

Ironically, the function does not get its name because it is heavy on one side but, instead, from the British mathematician and physicist, Oliver Heaviside.

5.4.2   The Derivative of $\delta(t)$

The derivatives of $\delta(t)$ are, themselves, very interesting functions with sifting properties of their own. To explore these functions, picture the first derivative of $\delta(t)$ as the limit of the derivatives of the Gaussian sequence functions:

$$\frac{d\delta(t)}{dt} = \lim_{n \to \infty} \frac{d\delta_n(t)}{dt} = \lim_{n \to \infty} \frac{d}{dt}\left[\frac{n}{\sqrt{\pi}}e^{-n^2 t^2}\right]. \tag{5.39}$$

Figure 5.15 depicts the limiting process.

The first derivative of $\delta(t)$, often written $\delta'(t)$, is called a "doublet" because of its opposing pair of spikes, which are infinitely high, infinitely narrow, and infinitely close together. Like $\delta(t)$, the doublet is rigorously defined by how it operates on other functions inside an integral. The sifting integral of the doublet is

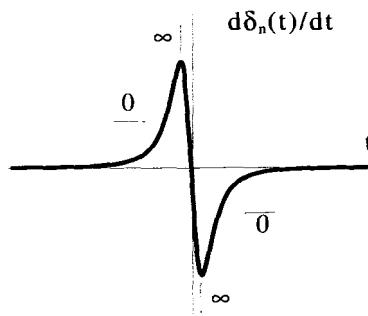$$\int_{t_-}^{t_+} dt\; f(t)\delta'(t), \tag{5.40}$$

$d\delta_n(t)/dt$

Figure 5.15   The Derivative of $\delta(t)$

where $f(t)$ is any continuous, well-behaved function. This integral can be evaluated using integration by parts. Recall that integration by parts is performed by identifying two functions, $u(t)$ and $v(t)$, and using them in the relation

$$\int_a^b d(u(t))v(t) = u(t)v(t)\Big|_{t=a}^{t=b} - \int_a^b d(v(t))u(t). \tag{5.41}$$

To evaluate Equation 5.40, let $v(t) = f(t)$ and $du = \delta'(t)\,dt$. Then $dv = (df/dt)\,dt$ and $u(t) = \delta(t)$ so that Equation 5.40 becomes

$$\int_{t_-}^{t_+} dt\, f(t)\delta'(t) = f(t)\delta(t)\Big|_{t=t_-}^{t=t_+} - \int_{t_-}^{t_+} dt\, \frac{df(t)}{dt}\delta(t). \tag{5.42}$$

If $t_- \neq 0$ and $t_+ \neq 0$, Equation 5.42 reduces to

$$\int_{t_-}^{t_-} dt\, f(t)\delta'(t) = -\int_{t_-}^{t_+} dt\, \frac{df(t)}{dt}\delta(t), \tag{5.43}$$

because $\delta(t)$ is zero for $t \neq 0$. This is now in the form of the standard sifting integral, so

$$\int_{t_-}^{t_+} dt\, f(t)\delta'(t) = \begin{cases} -f'(0) & t_- < 0 < t_+ \\ 0 & \text{otherwise} \end{cases}. \tag{5.44}$$

The doublet is the sifting function for the negative of the derivative.

This sifting property of the doublet, like the sifting property of $\delta(t)$, has a graphical interpretation. The integrated area under the doublet is zero, so if it is multiplied by a constant, the resulting integral is also zero. However, if the doublet is multiplied by a function that has different values to the right and left of center (i.e., a function with a nonzero derivative at that point), then one side of the integrand will have more area than the other. If the derivative of $f(t)$ is positive, the negative area of the doublet is scaled more than the positive area and the result of the sifting integration is negative. This situation is shown in Figure 5.16. If, on the other hand, the derivative of $f(t)$
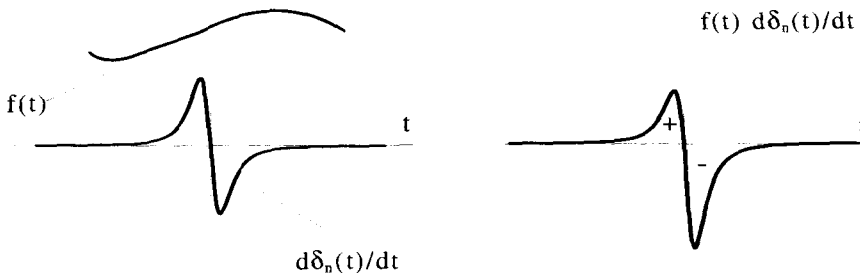

Figure 5.16   The Sifting Property of the Doublet

is negative, the positive area of the doublet is enhanced more than the negative area, and the sifting integration produces a positive result.

Similar arguments can be extended to higher derivatives of the δ-function. The appropriate sifting property is arrived at by applying integration by parts a number of times, until the integral is placed in the form of Equation 5.44.

5.5   SINGULAR DENSITY FUNCTIONS

One of the more common uses of the Dirac δ-function is to describe density functions for singular distributions. Typically, a distribution function describes a continuous, "per unit volume" quantity such as charge density, mass density, or number density. Singular distributions are not continuous, but describe distributions that are confined to sheets, lines, or points. The example of the density of a point mass was briefly described earlier in this chapter.

5.5.1   Point Mass Distributions

The simplest singular mass distribution describes a point of mass m_o located at the origin of a Cartesian coordinate system, as shown in Figure 5.4. The mass density function $\rho_m(x, y, z)$ that describes this distribution must have the units of mass per volume and must be zero everywhere, except at the origin. In addition, any volume integral of the density which includes the origin must give a total mass of m_o, while integrals that exclude the origin must give zero mass. The expression

$$\rho_m(x, y, z) = m_o \delta(x)\delta(y)\delta(z) \tag{5.45}$$

satisfies all of these requirements. Clearly it is zero unless x, y, and z are zero. Therefore integrating ρ_m over a volume that does not include the origin produces zero. Integrating over a volume that contains the origin results in

$$\int_V d\tau \, m_o \delta(x)\delta(y)\delta(z) = \int_{-\epsilon}^{+\epsilon} dx \int_{-\epsilon}^{+\epsilon} dy \int_{-\epsilon}^{+\epsilon} dz \, m_o \delta(x)\delta(y)\delta(z)$$
$$= m_o. \tag{5.46}$$

Because $\int dx \, \delta(x) = 1$, $\delta(x)$ has the inverse dimensions of its argument, or in this case, 1/length. Therefore, $m_o \delta(x)\delta(y)\delta(z)$ has the proper dimensions of mass per unit volume.

To shift the point mass to a different location than the origin, simply use shifted δ-functions:

$$\rho_m = m_o \delta(x - x_o)\delta(y - y_o)\delta(z - z_o). \tag{5.47}$$

This function has the proper dimensions of mass per unit volume, and ρ_m is zero except at the point (x_o, y_o, z_o). The integral over a volume V correctly produces zero mass if V does not include the point (x_o, y_o, z_o) and m_o if it does.

Consider the very same point, but now try to write the mass density in cylindrical coordinates, i.e., $\rho_m(\rho, \phi, z)$. In this system, the coordinates of the mass point are (ρ_o, ϕ_o, z_o) where

$$\rho_o = \sqrt{x_o^2 + y_o^2}$$

$$\phi_o = \tan^{-1}(y_o/x_o) \tag{5.48}$$

$$z_o = z_o.$$

A natural guess for $\rho_m(\rho, \phi, z)$ might be

$$m_o\delta(\rho - \rho_o)\delta(\phi - \phi_o)\delta(z - z_o). \tag{5.49}$$

This density clearly goes to zero, unless $\rho = \rho_o$, $\phi = \phi_o$ and $z = z_o$, as it should. However, because ϕ is dimensionless, Equation 5.49 has the dimensions of mass per unit area, which is not correct. Also, the integral over a volume V becomes

$$\int_V d\tau\, \rho_m = \int d\rho \int \rho\, d\phi \int dz\, m_o\delta(\rho - \rho_o)\delta(\phi - \phi_o)\delta(z - z_o). \tag{5.50}$$

This is zero if V does not contain the point (ρ_o, ϕ_o, z_o), but when V does contain the point, the result is $m_o\rho_o$, not m_o as required. This is because the $d\rho$ integration gives

$$\int_{\rho_o-\epsilon}^{\rho_o+\epsilon} d\rho\, \rho\delta(\rho - \rho_o) = \rho_o. \tag{5.51}$$

Therefore, Equation 5.49 is not the proper expression for the mass density in cylindrical coordinates, because it does not have the proper dimensions, and integration does not produce the total mass. From the discussion above, it is clear that the correct density function is

$$\rho_m = \frac{m_o}{\rho}\delta(\rho - \rho_o)\delta(\phi - \phi_o)\delta(z - z_o). \tag{5.52}$$

This example demonstrates an important point. While the point distributions in a Cartesian system can be determined very intuitively, a little more care must be used with non-Cartesian coordinates. In generalized curvilinear coordinates, where $d\tau = h_1 h_2 h_3 dq_1 dq_2 dq_3$, the expression for a point mass at (q_{1o}, q_{2o}, q_{3o}) is

$$\rho_m(q_1, q_2, q_3) = m_o\frac{\delta(q_1 - q_{1o})}{h_1}\frac{\delta(q_2 - q_{2o})}{h_2}\frac{\delta(q_3 - q_{3o})}{h_3}. \tag{5.53}$$

There is a common shorthand notation used for three-dimensional singular distributions. The three-dimensional Dirac δ-function is defined by

$$\delta^3(\bar{r} - \bar{r}_o) \equiv \frac{\delta(q_1 - q_{1o})}{h_1}\frac{\delta(q_2 - q_{2o})}{h_2}\frac{\delta(q_2 - q_{2o})}{h_2}. \tag{5.54}$$

In Cartesian coordinates, this is simply

$$\delta^3(\bar{\mathbf{r}} - \bar{\mathbf{r}}_o) = \delta(x - x_o)\delta(y - y_o)\delta(z - z_o). \qquad (5.55)$$

Using this notation, the mass density of point mass located at $\bar{\mathbf{r}}_o$ is, regardless of the coordinate system chosen, given by

$$\rho_m(\bar{\mathbf{r}}) = m_o\delta^3(\bar{\mathbf{r}} - \bar{\mathbf{r}}_o). \qquad (5.56)$$

If there are several points with mass m_i at position $\bar{\mathbf{r}}_i$, the density function becomes a sum of δ-functions:

$$\rho_m(\bar{\mathbf{r}}) = \sum_i m_i\delta^3(\bar{\mathbf{r}} - \bar{\mathbf{r}}_i). \qquad (5.57)$$

Integrating over a volume V gives

$$\int_V d\tau \sum_i m_i\delta^3(\bar{\mathbf{r}} - \bar{\mathbf{r}}_i), \qquad (5.58)$$

which, when evaluated, is simply the sum of all the masses enclosed in V.

5.5.2   Sheet Distributions

Imagine a two-dimensional planar sheet of uniform mass per unit area σ_o, located in the $z = z_o$ plane of a Cartesian system, as shown in Figure 5.17. The intuitive guess for the mass density ρ_m in Cartesian coordinates is

$$\rho_m(x, y, z) = \sigma_o\delta(z - z_o). \qquad (5.59)$$

The δ-function makes sure all the mass is in the $z = z_o$ plane. The dimensions are correct, because σ_o has the dimensions of mass per area, and the δ-function adds another 1/length, to give ρ_m the dimensions of mass per unit volume. The real check,
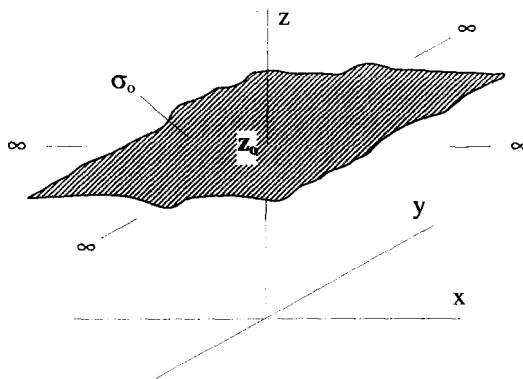

Figure 5.17   Simple Infinite Planar Sheet

however, is the volume integral. In order for Equation 5.59 to be correct, an integral of the surface mass density over some part of the sheet S must give the same total mass as a volume integral of ρ_m, over a volume V, which encloses S:

$$\int_S d\sigma \ \sigma_m(\bar{\mathbf{r}}) = \int_V d\tau \ \rho_m(\bar{\mathbf{r}}). \tag{5.60}$$

For the case described above, Equation 5.60 expands on both sides to

$$\int dx \int dy \ \sigma_o = \int dx \int dy \int dz \ \sigma_o \delta(z - z_o), \tag{5.61}$$

which is indeed true, because the range of the z integration includes the zero of the δ-function. Thus the assumption of Equation 5.59 was correct. Notice how the δ-function effectively converts the volume integral into a surface integral.

Unfortunately, things are not always this easy. The previous example was particularly simple, because the sheet was lying in the plane given by $z = z_o$. Now consider the same sheet positioned in the plane $y = x$, as shown in Figure 5.18. In this case the mass is only located where $y = x$, and the intuitive guess for the mass density is

$$\sigma_o \delta(y - x). \tag{5.62}$$

But our intuition is incorrect in this case. The surface integral on the LHS of Equation 5.60 expands to

$$\int d\sigma \ \sigma_m(\bar{r}) = \int ds \int dz \ \sigma_o, \tag{5.63}$$

where ds is the differential length on the surface in the xy-plane, as shown in Figure 5.19. Notice that

$$ds = \sqrt{(dx)^2 + (dy)^2}$$
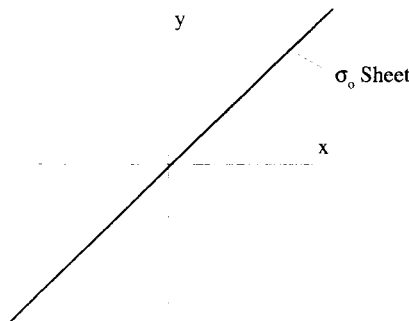$$= \sqrt{2} \, dx, \tag{5.64}$$

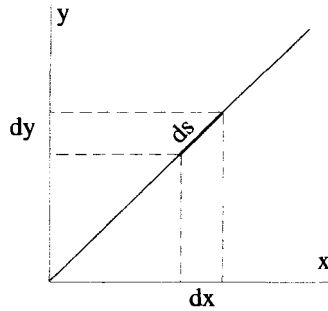**Figure 5.18** A Tilted Planar Mass Sheet

Figure 5.19 The Differential Length ds Along the Surface $x = y$ in the xy-Plane

where the last step follows because $dx = dy$ for this problem. Thus Equation 5.63 becomes

$$\int d\sigma \, \sigma_m(\bar{r}) = \sqrt{2} \int dx \int dz \, \sigma_o. \qquad (5.65)$$

If we use the expression in 5.62 for ρ_m, the volume integral on the RHS of Equation 5.60 is

$$\int d\tau \, \rho(\bar{r}) = \int dx \int dy \int dz \, \sigma_o \delta(y - x)$$

$$= \int dx \int dz \, \sigma_o, \qquad (5.66)$$

where the last step results from performing the integration over y. Comparing Equations 5.66 and 5.65, we see that they are off by a factor of $\sqrt{2}$!

Where does this discrepancy come from? The problem was our assumption in Equation 5.62 to use a δ-function in the form $\delta(y - x)$. We could have very well chosen $[?]\delta(y - x)$, where $[?]$ is some function that we need to determine. This still makes all the mass lie in the $y = x$ plane. The correct choice of $[?]$ is the one that makes both sides of Equation 5.60 equal. For example, when we make our "guess" for the distribution function for this example, we write

$$\rho_m(x, y, z) = [?]\sigma_o \delta(x - y). \qquad (5.67)$$

Then, when we evaluate the RHS of Equation 5.60, we get

$$\int d\tau \, \rho(\bar{r}) = \int dx \int dy \int dz \, \sigma_o [?]\delta(y - x)$$

$$= \int dx \int dz \, [?]\sigma_o. \qquad (5.68)$$

In this case, we already showed that the value of $[?]$ is simply the constant $\sqrt{2}$. In more complicated problems, $[?]$ can be a function of the coordinates. This can happen

if either the mass per unit area of the sheet is not constant, or if the sheet is not flat. You will see some examples of this in the next section and in the problems at the end of this chapter.

5.5.3   Line Distributions

As a final example of singular density functions, consider the mass per unit volume of a one-dimensional wire of uniform mass per unit length λ_o. The wire is bent to follow the parabola $y = Cx^2$ in the $z = 0$ plane, as shown in Figure 5.20. The factor C is a constant, which has units of $1/\text{length}$. We will follow the same procedure in constructing the mass density for this wire as we did for the previous example. In this case, the volume integral of the mass density must collapse to a line integral along the wire

$$\int_V d\tau \, \rho_m(\bar{\mathbf{r}}) = \int_L ds \, \lambda_m(s). \qquad (5.69)$$

In this equation, s is a variable which indicates parametrically where we are on the wire, and $\lambda_m(s)$ is the mass per unit length of the wire at the position s.

Because all the mass must lie on the wire, we write the mass density function as

$$\rho_m(x, y, z) = [?]\lambda_o \delta(y - Cx^2)\delta(z). \qquad (5.70)$$

Here we have used two δ-functions. The $\delta(z)$ term ensures all the mass lies in the $z = 0$ plane, while $\delta(y - Cx^2)$ makes the mass lie along the parabola. As before, we include an unknown factor of $[?]$, which we will have to determine using Equation 5.69.

In terms of Cartesian coordinates, the general expression for the differential arc length ds is

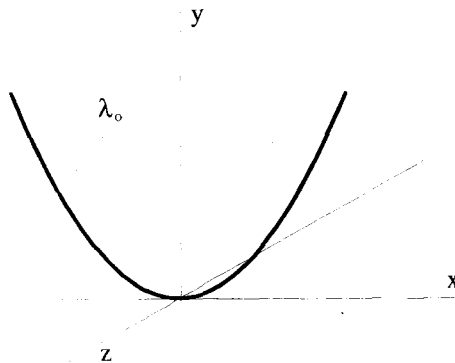$$ds = \sqrt{(dx)^2 + (dy)^2 + (dz)^2}. \qquad (5.71)$$

Figure 5.20   Parabolic Line Distribution

Along the wire, $z = 0$ and $y = Cx^2$, so that

$$ds = \sqrt{1 + 4C^2x^2}\, dx. \tag{5.72}$$

Using Equation 5.70 and the fact that $d\tau = dx\, dy\, dz$, Equation 5.69 becomes

$$\int dx \int dy \int dz\, [?]\lambda_o \delta(y - Cx^2)\delta(z) = \int dx\, \sqrt{1 + 4C^2x^2}\, \lambda_o. \tag{5.73}$$

Let's concentrate on the LHS of this equation. The integral over z is easy, because

$$\int dz\, \delta(z) = 1. \tag{5.74}$$

Also, when we do the integral over y, x is held constant, and only one value of y makes the argument of the δ-function vanish, so we have

$$\int dy\, \delta(y - Cx^2) = 1. \tag{5.75}$$

Therefore, Equation 5.73 becomes

$$\int dx\, [?]\lambda_o = \int dx\, \sqrt{1 + 4C^2x^2}\, \lambda_o, \tag{5.76}$$

and the value of [?] clearly must be

$$[?] = \sqrt{1 + 4C^2x^2}. \tag{5.77}$$

Notice in this case that [?] is a function of position. The mass density for the wire is given by

$$\rho_m(x, y, z) = \sqrt{1 + 4C^2x^2}\, \lambda_o \delta(y - Cx^2)\delta(z). \tag{5.78}$$

When we converted the volume integral in Equation 5.73 to a line integral, it was easier to perform the dy integration before the dx integration. This was because when we integrated over y, holding x fixed, the only value of y which made the argument of the δ-function zero was $y = Cx^2$. Another way to look at this is that there is a one-to-one relationship between dx and ds, as shown in Figure 5.21.

If instead, we performed the x integration first, holding the value of y fixed, there are *two* values of x which zero the δ-function argument. In this case, the integral becomes

$$\int dx\, \delta(y - Cx^2) = \frac{1}{C} \int dx\, \delta\left(\left[x - \sqrt{y/C}\,\right]\left[x + \sqrt{y/C}\,\right]\right)$$

$$= \frac{1}{\sqrt{4Cy}} \int dx\, \left[\delta\left(x - \sqrt{y/C}\,\right) + \delta\left(x + \sqrt{y/C}\,\right)\right]. \tag{5.79}$$
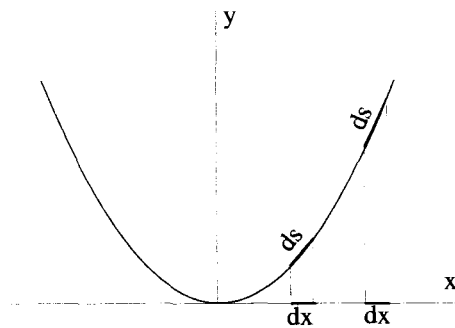
Figure 5.21   The Relation Between dx and ds Along the Parabola
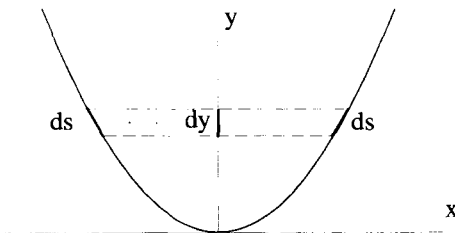
Figure 5.22   The Relation Between dy and ds Along the Parabola

Looking at Figure 5.22 shows what is going on here. There is not a one-to-one relationship between dy and ds. Of course, evaluating the integral in this way produces exactly the same result for $\rho_m(\bar{r})$, as you will prove in one of the exercises of this chapter.

5.6   THE INFINITESIMAL ELECTRIC DIPOLE

The example of the infinitesimal electric dipole is one of the more interesting applications of the Dirac δ-function and makes use of many of its properties.

5.6.1   Moments of a Charge Distribution

In electromagnetism, the distribution of charge density in space $\rho_c(\bar{r})$ can be expanded, in what is generally called a multipole expansion, into a sum of its *moments*. These moments are useful for approximating the potential fields associated with complicated charge distributions in the far field limit (that is, far away from the charges). Each moment is generated by calculating a different volume integral of the charge distribution over all space. Because our goal is not to derive the mathematics of multipole expansions, but rather to demonstrate the use of the Dirac δ-functions, the multipole expansion results are stated here without proof. Derivations can be found

in most any intermediate or advanced book on electromagnetism, such as Jackson's *Classical Electrodynamics*.

The lowest term in the expansion is a scalar called the monopole moment. It is just the total charge of the distribution and is determined by calculating the volume integral of ρ_c:

$$Q = \int_{\text{All space}} d\tau \, \rho_c(\bar{\mathbf{r}}). \tag{5.80}$$

The next highest moment is a vector quantity called the dipole moment, which is generated from the volume integral of the charge density times the position vector:

$$\bar{\mathbf{P}} = \int_{\text{All space}} d\tau \, \bar{\mathbf{r}}\rho_c(\bar{\mathbf{r}}). \tag{5.81}$$

The next moment, referred to as the quadrapole moment, is a tensor quantity generated by the integral

$$\bar{\bar{\mathbf{Q}}} = \int_{\text{All space}} d\tau \left(3\bar{\mathbf{r}}\,\bar{\mathbf{r}} - |\bar{\mathbf{r}}|^2\bar{\bar{\mathbf{1}}}\right)\rho_c(\bar{\mathbf{r}}). \tag{5.82}$$

In this equation, the quantity $\bar{\mathbf{r}}\,\bar{\mathbf{r}}$ is a dyad, and $\bar{\bar{\mathbf{1}}}$ is the identity tensor. There are an infinite number of higher-order moments beyond these three, but they are used less frequently, usually only in cases where the first three moments are zero.

Far away from the charges, the electric potential can be approximated by summing the contributions from each of the moments. The potential field Φ due to the first few moments is

$$\Phi(\bar{\mathbf{r}}) = \frac{1}{4\pi\epsilon_o}\left[\frac{Q}{r} + \frac{\bar{\mathbf{r}} \cdot \bar{\mathbf{P}}}{r^3} + \frac{\bar{\mathbf{r}} \cdot \bar{\bar{\mathbf{Q}}} \cdot \bar{\mathbf{r}}}{2r^5} + \cdots\right]. \tag{5.83}$$

It is quite useful to know what charge distributions generate just a single term in this expansion, and what potentials and electric fields are associated with them. For example, what charge distribution has just a dipole term (that is, $\bar{\mathbf{P}} \neq 0$) while all other terms are zero ($Q = 0$, $\bar{\bar{\mathbf{Q}}} = 0$, etc.). The Dirac δ-function turns out to be quite useful in describing these particular distributions.

5.6.2 The Electric Monopole

The distribution that generates just the Q/r term in Equation 5.83 is called the electric monopole. As you may have suspected, it is simply the distribution of a point charge at the origin:

$$\rho_{\text{mono}} = q_o\delta^3(\bar{\mathbf{r}}). \tag{5.84}$$

Its monopole moment,

$$Q = \int_{\text{All space}} d\tau \, q_o \delta^3(\bar{\mathbf{r}}) = q_o, \tag{5.85}$$

is simply equal to the total charge. The dipole, quadrapole, and higher moments of this distribution are all zero:

$$\bar{\mathbf{P}} = \int_{\text{All space}} d\tau \, q_o \bar{\mathbf{r}} \, \delta^3(\bar{\mathbf{r}}) = 0 \tag{5.86}$$

$$\overline{\overline{\mathbf{Q}}} = \int_{\text{All space}} d\tau \left(3\bar{\mathbf{r}}\,\bar{\mathbf{r}} - r^2 \overline{\overline{\mathbf{I}}} \right) \delta^3(\bar{\mathbf{r}}) = 0. \tag{5.87}$$

The electric field of a monopole obeys Coulomb's law:

$$\bar{\mathbf{E}} = \frac{1}{4\pi\epsilon_o} \frac{q_o \hat{\mathbf{e}}_r}{r^2}. \tag{5.88}$$

5.6.3 The Electric Dipole

An electric dipole consists of two equal point charges, of opposite sign, separated by some finite distance d_o, as shown in Figure 5.23. The charge density of this system, expressed using Dirac δ-functions, is

$$\rho_{\text{dpl}} = q_o \left[\delta^3 \left(\bar{\mathbf{r}} - \frac{d_o}{2}\hat{\mathbf{e}}_x \right) - \delta^3 \left(\bar{\mathbf{r}} + \frac{d_o}{2}\hat{\mathbf{e}}_x \right) \right], \tag{5.89}$$

where in this case, the dipole is oriented along the x-axis.

You might be tempted to believe that this distribution has only a dipole moment. Indeed both Q and $\overline{\overline{\mathbf{Q}}}$ are zero, while the dipole moment is given by

$$\begin{aligned}
\bar{\mathbf{P}} &= \int d\tau \, q_o \bar{\mathbf{r}} \left[\delta^3 \left(\bar{\mathbf{r}} - \frac{d_o}{2}\hat{\mathbf{e}}_x \right) - \delta^3 \left(\bar{\mathbf{r}} + \frac{d_o}{2}\hat{\mathbf{e}}_x \right) \right] \\
&= \frac{q_o d_o}{2}\hat{\mathbf{e}}_x + \frac{q_o d_o}{2}\hat{\mathbf{e}}_x \\
&= q_o d_o \hat{\mathbf{e}}_x.
\end{aligned} \tag{5.90}$$

The higher-order moments for the dipole distribution, however, do not vanish.
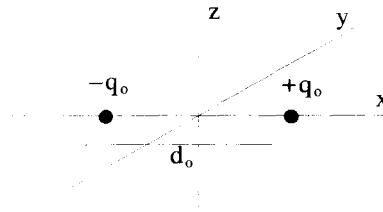

Figure 5.23 The Electric Dipole

The "ideal" dipole refers to a charge distribution that has only a dipole moment, and no other. It is the limiting case of the "physical" dipole, described above. In this limit, the distance between the charges becomes vanishingly small, while the net dipole moment $\overline{\mathbf{P}}$ is held constant. In other words, $d_o \to 0$ and $q_o \to \infty$ such that $d_o q_o \equiv p_o$ is held constant. The charge distribution for this situation is

$$\rho_{\text{ideal}} = \lim_{d_o \to 0} \frac{p_o}{d_o} \left[\delta^3 \left(\overline{\mathbf{r}} - \frac{d_o}{2} \hat{\mathbf{e}}_x \right) - \delta^3 \left(\overline{\mathbf{r}} + \frac{d_o}{2} \hat{\mathbf{e}}_x \right) \right]. \tag{5.91}$$

Expansion of the δ-functions in terms of Cartesian coordinates gives

$$\rho_{\text{ideal}} = p_o \delta(y) \delta(z) \lim_{d_o \to 0} \frac{1}{d_o} \left[\delta \left(x - \frac{d_o}{2} \right) - \delta \left(x + \frac{d_o}{2} \right) \right]. \tag{5.92}$$

Notice that this limit is just the definition of a derivative:

$$\lim_{d_o \to 0} \frac{1}{d_o} \left[\delta \left(x - \frac{d_o}{2} \right) - \delta \left(x + \frac{d_o}{2} \right) \right] = -\frac{d\delta(x)}{dx}. \tag{5.93}$$

This means the charge distribution of an ideal dipole, with a dipole moment of magnitude p_o oriented along the x axis, can be written

$$\rho_{\text{ideal}} = -p_o \frac{d\delta(x)}{dx} \delta(y) \delta(z). \tag{5.94}$$

The electric field from the ideal dipole (and also from a physical dipole in the far field limit) is the solution of

$$\overline{\nabla} \cdot \overline{\mathbf{E}} = \frac{\rho(\overline{\mathbf{r}})}{\epsilon_o} = -\frac{p_o}{\epsilon_o} \frac{d\delta(x)}{dx} \delta(y) \delta(z). \tag{5.95}$$

But we already know the solution for the electric monopole is

$$\overline{\nabla} \cdot \left\{ \frac{1}{4\pi\epsilon_o} \frac{q_o \hat{\mathbf{e}}_r}{r^2} \right\} = \frac{q_o}{\epsilon_o} \delta(x) \delta(y) \delta(z). \tag{5.96}$$

Operating on both sides of Equation 5.96 with $[-d_o \partial / \partial x]$, shows the electric field in Equation 5.95 obeys

$$\overline{\mathbf{E}} = -d_o \frac{\partial}{\partial x} \left\{ \frac{1}{4\pi\epsilon_o} \frac{q_o \hat{\mathbf{e}}_r}{r^2} \right\}. \tag{5.97}$$

Taking the derivative gives the result:

$$\overline{\mathbf{E}} = \frac{p_o}{4\pi\epsilon_o r^3} \left\{ \frac{3x}{r} \hat{\mathbf{e}}_r - \hat{\mathbf{e}}_x \right\}. \tag{5.98}$$

This is the solution for a dipole oriented along the x-axis. In general, the dipole is a vector $\overline{\mathbf{P}}$, which can be oriented in any direction, and Equation 5.98 generalizes to

$$\overline{\mathbf{E}} = \frac{1}{4\pi\epsilon_o r^3} \left\{ 3(\overline{\mathbf{P}} \cdot \hat{\mathbf{e}}_r)\hat{\mathbf{e}}_r - \overline{\mathbf{P}} \right\}. \tag{5.99}$$

Notice that the magnitude of this field drops off faster with r than the field of the monopole. The higher the order of the moment, the faster its field decays with distance.

5.6.4   Fields from Higher-Order Moments

The technique described above can be used to find the electric field and charge distributions for the ideal electric quadrapole, as well as for all the higher moments. The charge distributions can be constructed using δ-functions, and the fields can be obtained by taking various derivatives of the monopole field. You can practice this technique with the quadrapole moment in an exercise at the end of this chapter.

5.7   RIEMANN INTEGRATION AND THE DIRAC δ-FUNCTION

The δ-function provides a useful, conceptual technique for viewing integration which will become important when we discuss Green's functions in a later chapter. From Equation 5.14, the sifting integral definition of the δ-function, any continuous function $f(y)$ can be written

$$f(y) = \int_{-\infty}^{+\infty} dx \, \delta(x - y) f(x). \tag{5.100}$$

The Riemann definition of integration says that an integral can be viewed as the limit of a discrete sum of rectangles,

$$\int_{-\infty}^{+\infty} dx \, g(x) = \lim_{\Delta x \to 0} \sum_{n=-\infty}^{n=+\infty} g(n\Delta x)\Delta x, \tag{5.101}$$

where Δx is the width of the rectangular blocks which subdivide the area being integrated, and n is an integer which indexes each rectangle. The limiting process increases the number of rectangles, so for well-behaved functions, the approximation becomes more and more accurate as $\Delta x \to 0$. The Riemann definition is pictured graphically in Figures 5.24(a) and (b).

Using the Riemann definition, Equation 5.100 becomes

$$f(y) = \int_{-\infty}^{+\infty} dx \, f(x)\delta(x - y)$$

$$= \lim_{\Delta x \to 0} \sum_{n=-\infty}^{n=+\infty} f(n\Delta x)\delta(n\Delta x - y)\Delta x. \tag{5.102}$$
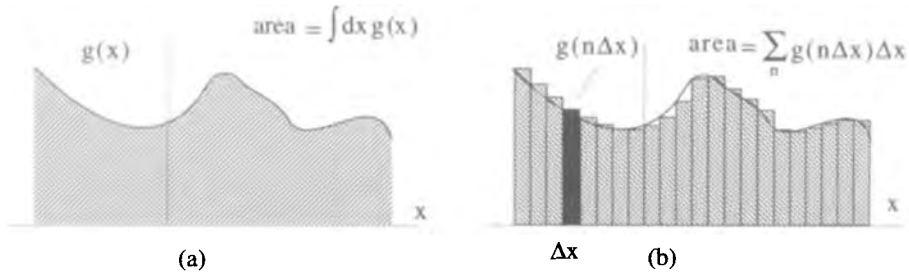
Figure 5.24   Discrete Sum Representation of an Integral
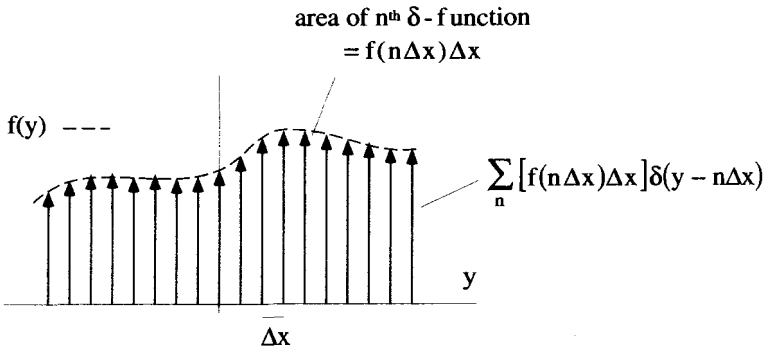
Figure 5.25   The Construction of $f(y)$ from an Infinite Sum of δ-Functions

Equation 5.102 makes a very interesting statement: Any continuous function can be viewed as the sum of an infinite number of δ-functions. Both $f(y)$ and the RHS of Equation 5.102 are functions of y, and are plotted in Figure 5.25 for finite Δx. In this figure, the δ-functions are located at $y = n\Delta x$ and have an area $f(n\Delta x)\Delta x$. The function $f(y)$ is generated by this sum of δ-functions as $\Delta x \to 0$, i.e., as the spacing between the δ-functions and their areas go to zero. Thus an infinite number of infinitesimal area δ-functions, spaced arbitrarily close together, combine to form the continuous function $f(y)$!

EXERCISES FOR CHAPTER 5

1. Simplify the integral

$$\int_{-\infty}^{\infty} dx\, f(x)\delta(-ax + b),$$

where a and b are real, positive constants.

2. Perform these integrations which involve the Dirac δ-function:

   i. $\int_{-3}^{0} dx \, \delta(x - 1)$.

   ii. $\int_{-\infty}^{\infty} dx \, (x^2 + 3)\delta(x - 5)$.

   iii. $\int_{-5}^{5} dx \, x \, \delta(x^2 - 5)$.

   iv. $\int_{-5}^{5} dx \, x \, \delta(x^2 + 5)$.

   v. $\int_{0}^{2\pi} dx \, \delta(\cos x)$.

   vi. $\int_{\pi/4}^{3\pi/4} dx \, x^2 \, \delta(\cos x)$.

   vii. $\int_{-\infty}^{\infty} dx \, x^2 \, \frac{d\delta(x)}{dx}$.

   viii. $\int_{-10}^{10} dx \, (x^2 + 3) \left[\frac{d\delta(x - 5)}{dx} \right]$.

3. Determine the integral properties of the "triplet" $d^2\delta(t)/dt^2$ by evaluating the integrals

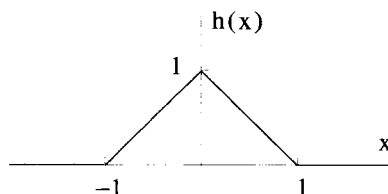   **i.** $\int_{-\infty}^{\infty} dt \left(\frac{d^2\delta(t)}{dt^2} \right)$.

   ii. $\int_{-\infty}^{\infty} dt \, f(t) \left(\frac{d^2\delta(t - t_o)}{dt^2} \right)$.

4. The function $h(x)$ is generated from the function $g(x)$ by the integral

$$h(x) = \int_{-\infty}^{\infty} dx' \, g(x') \frac{d\delta(x' - x)}{dx'}.$$

If $h(x)$ is the triangular pulse shown below, find and plot $g(x)$.

5. Given that

$$f(x) = \begin{cases} 0 & x < 0 \\ e^x - 1 & x > 0 \end{cases},$$

find and plot the first and second derivatives of $f(x)$.

6. An ideal impulse moving in space and time can be described by the function,

$$f(x, t) = I_o \delta(x - v_o t),$$

where v_o is a constant. Make a three-dimensional plot for $f(x, t)$ vs. x and t to show how this impulse propagates. What are the dimensions of v_o? How does the plot change if

$$f(x, t) = I_o \delta(x - a_o t^2 / 2),$$

where a_o is a constant? What are the dimensions of a_o?

7. Consider the sinc function sequence:

$$\delta_n(t) = \sin(n\pi)/(\pi x).$$

(a) On the same graph, plot three of these functions with $n = 1$, $n = 10$, and $n = 100$.

(b) Prove that the limit of the sinc sequence functions acts like a δ-function by deriving the relation:

$$\int_{t_-}^{t_+} dt \left[\lim_{n \to \infty} \frac{\sin nt}{\pi t} \right] f(t) = \begin{cases} f(0) & t_- < 0 < t_+ \\ 0 & \text{otherwise} \end{cases}. \qquad (5.103)$$

As a first step, try making the substitution $y = nt$. You will need to use the identity:

$$\int_{-\infty}^{\infty} dx \frac{\sin x}{x} = \pi.$$

8. The function $\delta(\cos x)$ can be written as a sum of Dirac δ-functions

$$\delta(\cos x) = \sum_n a_n \delta(x - x_n).$$

Find the range for n and the values for the a_n and the x_n.

9. A single point charge q_o is located at $(1, 1, 0)$ in a Cartesian coordinate system, so that its charge density can be expressed as

$$\rho_c(x, y, z) = q_o \, \delta(x - 1) \, \delta(y - 1) \, \delta(z).$$

    (a) What is $\rho_c(\rho, \theta, z)$, its charge density in cylindrical coordinates?

    (b) What is $\rho_c(r, \theta, \phi)$, its charge density in spherical coordinates?

10. An infinitely long, one-dimensional wire of mass per unit length λ_o is bent to follow the curve $y = A \cosh(Bx)$, as shown below.
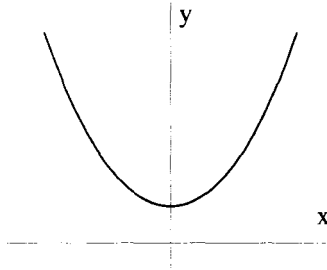
Find an expression for the mass per unit volume $\rho(x, y, z)$. Express your answer two ways:

(a) As the product of two δ-functions.

(b) As a sum of two terms, each the product of two δ-functions.

Be sure to check the dimensions of your answers.

11. An infinitely long, one-dimensional wire of mass per unit length λ_o is bent to follow the line formed by the intersection of the surface $x = y$ with the surface $y = z^2$, as shown in the figure below. Find an expression for $\rho_m(x, y, z)$, the mass per unit volume of the wire.
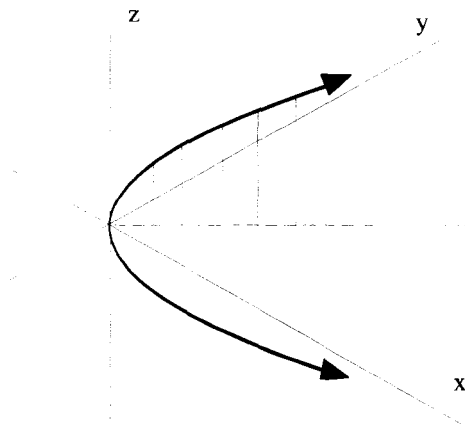
12. An infinitely long, one-dimensional wire with a constant mass per unit length λ_o is bent to follow the curve $y = \sin x$ in the $z = 0$ plane.
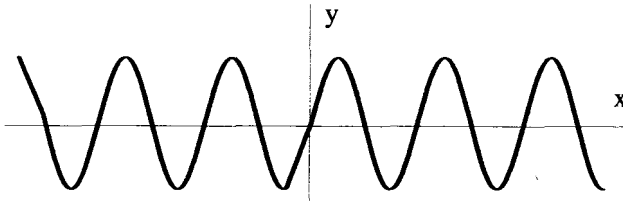
Determine the mass density $\rho_m(x, y, z)$ that describes this mass distribution.

13. A wire of mass per unit length λ_o is bent to follow the shape of a closed ellipse that lies in the xy-plane and is given by the expression

$$x^2 + 2y^2 = 4.$$

Express $\rho_m(x, y, z)$, the mass per unit volume of this object, using Dirac δ-functions. Show that your expression has the proper dimensions. There is more than one way to express the answer to this problem. Identify the most compact form.

14. An infinite, one-dimensional bar of mass per unit length λ_o lies along the line $y = m_o x$ in the $z = 0$ plane.
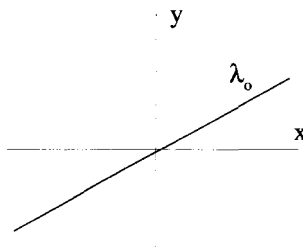
(a) Determine the mass per unit volume $\rho(x, y, z)$ of this bar.

(b) Now consider the situation where the bar is rotating about the z-axis at a constant angular velocity ω_o so that the angle the bar makes with respect to the x-axis is given by $\theta = \omega_o t$, as shown below. Find an expression for the time-dependent mass density $\rho(x, y, z, t)$.
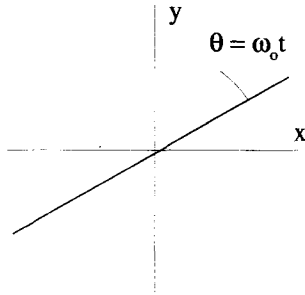
15. A charge Q_o is evenly distributed along the x-axis from $x = -L_o/2$ to $x = L_o/2$, as shown below.

   (a) Using the Heaviside step function, what is the charge density $\rho_c(x, y, z)$, expressed using Cartesian coordinates?

   (b) What is this charge density in cylindrical coordinates?
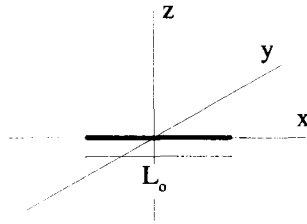
16. Using δ-functions and the Heaviside step function, express the charge density $\rho_c(\bar{r})$ of a uniformly charged cylindrical shell of radius r_o and length L_o. The total charge on the surface of the shell is Q_o.

17. An infinite, two-dimensional sheet with mass per unit area σ_o is bent to follow the surface $y = x^3$.

   (a) Make a plot of the curve made by the intersection of this sheet and the plane given by $z = 0$.

   (b) Determine the mass per unit volume $\rho_m(x, y, z)$.

18. Express the mass density $\rho_m(\rho, \theta, z)$ of a conical surface that is formed by cutting a pie-shaped piece from an infinite, uniform two-dimensional sheet of mass per unit area σ_o and joining the cut edges. The conical surface that results lies on the surface $\rho = a_o z$ where (ρ, θ, z) are the standard cylindrical coordinates.

19. An infinite, two-dimensional sheet with mass per unit area σ_o is bent to follow the surface $xy = 1$ in a Cartesian coordinate system.

   (a) Using the hyperbolic coordinates developed in Exercise 13 of Chapter 3, express the mass density $\rho_m(u, v, z)$ for this sheet.

   (b) Using the equations relating the coordinates, convert your answer to part (a) above to Cartesian coordinates.

   (c) Now, working from scratch in a Cartesian system, obtain $\rho_m(x, y, z)$ by requiring that this density function take the volume integral over all space to a surface integral over the hyperbolic surface.

20. Express the mass density $\rho_m(\bar{r})$ for a spherical sheet of radius r_o, with constant mass per unit area σ_o.

21. A dipole electric field is generated outside the surface of a sphere, if the charge per unit area on the surface of that sphere is distributed proportionally to $\cos(\theta)$. If the sphere has a radius r_o and there is a total charge of $+Q_o$ on the upper hemisphere and $-Q_o$ on the lower hemisphere, what is the expression for the charge density $\rho_c(r, \theta, \phi)$ in spherical coordinates?
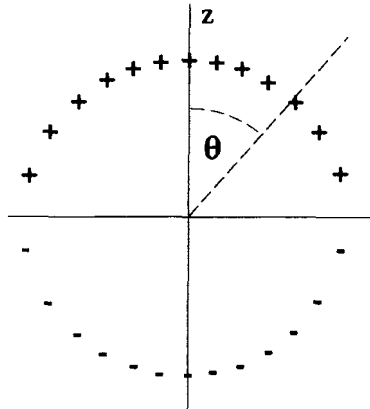
22. In a two-dimensional Cartesian coordinate system, the mass density $\rho_m(\bar{r})$ of a pair of point masses is given by

$$\rho_m(x_1, x_2) = m_o\delta(x_1 + 1)\delta(x_2) + m_o\delta(x_1 - 1)\delta(x_2).$$

Evaluate the integrals:

   i. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx_1 dx_2\, \rho_m(x_1, x_2).$

   ii. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx_1 dx_2\, \bar{r}\, \rho_m(x_1, x_2).$

iii. $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx_1 dx_2 \ [\bar{\mathbf{r}} \cdot \bar{\mathbf{r}}] \ \rho_m(x_1, x_2).$

iv. $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx_1 dx_2 \ [\bar{\mathbf{r}} \ \bar{\mathbf{r}}] \ \rho_m(x_1, x_2).$

23. Prove that the monopole and quadrapole moments of any dipole ("physical" or "ideal") are zero.

24. A quadrapole charge distribution consists of four point charges in the $x_1 x_2$-plane as shown below.
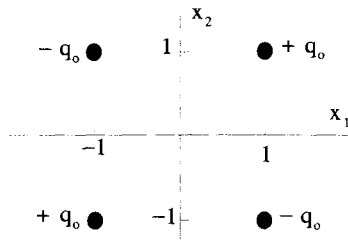
(a) Using Dirac δ-functions express the charge density, $\rho_c(x_1, x_2, x_3)$, of this distribution.

(b) The quadrapole moment of this charge distribution is a second rank tensor given by

$$\overline{\overline{\mathbf{Q}}} = Q_{ij} \ \hat{\mathbf{e}}_i \hat{\mathbf{e}}_j.$$

The elements of the quadrapole tensor are given by the general expression

$$Q_{ij} = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \int_{-\infty}^{\infty} dx_3 \ \rho_c(x_1, x_2, x_3) \left[3x_i x_j - (x_k x_k)\delta_{ij}\right],$$

where δ_{ij} is the Kronecker delta. In particular,

$$Q_{22} = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \int_{-\infty}^{\infty} dx_3 \ \rho_c(x_1, x_2, x_3) \left[3x_2 x_2 - (x_1^2 + x_2^2 + x_3^2)\right].$$

Evaluate all the elements of the quadrapole tensor for the charge distribution shown in the figure above.

(c) Does this charge distribution have a dipole moment?

(d) Find the coordinate system in which this quadrapole tensor is diagonal. Express the elements of $\overline{\overline{\mathbf{Q}}}$ in this system.

25. An ideal quadrapole has a charge density $\rho_c(x, y, y)$ that is zero everywhere except at the origin. It has zero total charge, zero dipole moment, and a nonzero quadrapole moment.

(a) Show that if $\rho_c(x, y, z)$ has the form

$$\rho_c(x, y, z) = [?]\frac{d\delta(x)}{dx}\frac{d\delta(y)}{dy}\delta(z),$$

it satisfies the above requirements for an ideal quadrapole. Evaluate [?] so that the elements of the quadrapole moment tensor for this distribution are the same as the quadrapole elements in Exercise 24.

(b) Determine the electric field produced by this ideal quadrapole.

6

INTRODUCTION TO
COMPLEX VARIABLES

One of the most useful mathematical skills to develop for solving physics and en-
gineering problems is the ability to move with ease in the complex plane. For that
reason, two chapters of this book are devoted to the subject of complex variable
theory. This first chapter is primarily restricted to the fundamentals of complex vari-
ables and single-valued complex functions. Some knowledge of complex numbers is
assumed.

6.1 A COMPLEX NUMBER REFRESHER

Complex numbers are based on the imaginary number "i", which is defined as the
positive square root of -1:

$$i \equiv \sqrt{-1}. \tag{6.1}$$

In general, a complex number $\underline{z}$ has a real and an imaginary part,

$$\underline{z} = x + iy, \tag{6.2}$$

where x and y are both real numbers. If y is zero for a particular complex number,
that number is called pure real. If x is zero, the number is pure imaginary.

6.1.1 The Complex Plane

A complex number can be plotted as a point in the *complex plane*. This plane is
simply a two-dimensional orthogonal coordinate system, where the real part of the
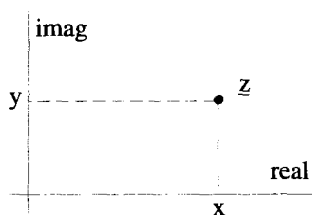
Figure 6.1   A Plot of the Complex Number $\underline{z} = x + iy$ in the Complex Plane

complex number is plotted on the horizontal axis and the imaginary part is plotted on the vertical axis. A plot of the complex number $\underline{z} = x + iy$ is shown in Figure 6.1. The *magnitude* of the complex number $\underline{z} = x + iy$ is

$$|\underline{z}| = \sqrt{x^2 + y^2}, \tag{6.3}$$

which is simply the distance from the origin to the point on the complex plane.

6.1.2   Complex Conjugates

The conjugate of a complex number can be viewed as the reflection of the number through the real axis of the complex plane. If $\underline{z} = x + iy$, its complex conjugate is defined as

$$\underline{z}^* \equiv x - iy. \tag{6.4}$$

The complex conjugate is useful for determining the magnitude of a complex number:

$$\underline{z}\,\underline{z}^* = (x + iy)(x - iy) = x^2 + y^2 = |\underline{z}|^2. \tag{6.5}$$

In addition, the real and imaginary parts of a complex variable can be isolated using the pair of expressions:

$$\frac{\underline{z} + \underline{z}^*}{2} = \frac{(x + iy) + (x - iy)}{2} = x \tag{6.6}$$

$$\frac{\underline{z} - \underline{z}^*}{2i} = \frac{(x + iy) - (x - iy)}{2i} = y. \tag{6.7}$$

6.1.3   The Exponential Function and Polar Representation

There is another representation of complex quantities which follows from Euler's equation,

$$e^{i\theta} = \cos\theta + i\sin\theta. \tag{6.8}$$

To understand this expression, consider the Taylor series expansion of e^x around zero:

$$e^x = e^x\big|_{x=0} + \frac{d(e^x)}{dx}\bigg|_{x=0} x + \frac{1}{2!}\frac{d^2(e^x)}{dx^2}\bigg|_{x=0} x^2 + \frac{1}{3!}\frac{d^3(e^x)}{dx^3}\bigg|_{x=0} x^3 \cdots$$

$$= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots. \tag{6.9}$$

Now we make the assumption that this Taylor series is valid for all numbers *including complex quantities*. Actually, we *define* the complex exponential function such that it is equivalent to this series:

$$e^{\underline{z}} \equiv 1 + \underline{z} + \frac{\underline{z}^2}{2!} + \frac{\underline{z}^3}{3!} + \frac{\underline{z}^4}{4!} + \cdots. \tag{6.10}$$

Now let $\underline{z} = i\theta$ to obtain

$$e^{i\theta} = 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \cdots. \tag{6.11}$$

Breaking this into real and imaginary parts gives

$$e^{i\theta} = \left[1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots\right] + i\left[\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots\right]. \tag{6.12}$$

The first bracketed term in Equation 6.12 is the Taylor series expansion for $\cos\theta$, and the second is the expansion for $\sin\theta$. Thus Euler's equation is proven.

Using this result, the complex variable $\underline{z} = x + iy$ can be written in the *polar representation*,

$$\underline{z} = re^{i\theta}$$

$$= r\cos\theta + ir\sin\theta, \tag{6.13}$$

where the new variables r and θ are defined by

$$r = |\underline{z}| = \sqrt{x^2 + y^2} \tag{6.14}$$

$$\theta = \tan^{-1} y/x. \tag{6.15}$$

The relationship between the two sets of variables is shown in Figure 6.2.
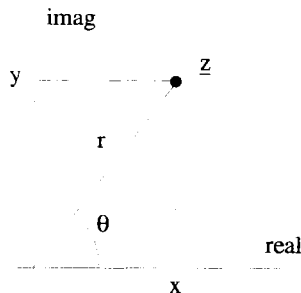
Figure 6.2   The Polar Variables

6.2 FUNCTIONS OF A COMPLEX VARIABLE

Just as with real variables, complex variables can be related to one another using functions. Imagine two complex variables, $\underline{w}$ and $\underline{z}$. A general function relating these two quantities can be written as

$$\underline{w} = \underline{w}(\underline{z}). \tag{6.16}$$

If $\underline{w} = u + iv$ and $\underline{z} = x + iy$, then we can rewrite this relationship as

$$\underline{w}(\underline{z}) = u(x, y) + iv(x, y), \tag{6.17}$$

where $u(x, y)$ and $v(x, y)$ are both real functions of real variables.

As a simple example of a complex function, consider the relationship

$$\underline{w}(\underline{z}) = \underline{z}^2. \tag{6.18}$$

Expanding $\underline{z} = x + iy$ gives

$$\underline{w}(\underline{z}) = (x^2 - y^2) + 2ixy. \tag{6.19}$$

The functions $u(x, y)$ and $v(x, y)$ from Equation 6.17 are easily identified as

$$u(x, y) = x^2 - y^2 \tag{6.20}$$

$$v(x, y) = 2xy. \tag{6.21}$$

6.2.1 Visualization of Complex Functions

A relationship between $\underline{w}$ and $\underline{z}$ can be visualized in two ways. The first interpretation is as a mapping of points from the complex $\underline{z}$-plane to the complex $\underline{w}$-plane. The function $\underline{w} = \underline{z}^2$ maps the point $\underline{z} = 2i$ into the point $\underline{w} = -4$ as shown in Figure 6.3. Another point $\underline{z} = 1 + i$ maps to $\underline{w} = 2i$. By mapping the entire z-plane onto the w-plane in this way, we can obtain a qualitative picture of the function.

A second, more abstract visualization is simply to plot the two functions $u(x, y)$ and $v(x, y)$ as the real and imaginary components of $\underline{w}$. This method is shown in Figure 6.4 for the function $\underline{w} = \sin \underline{z}$.
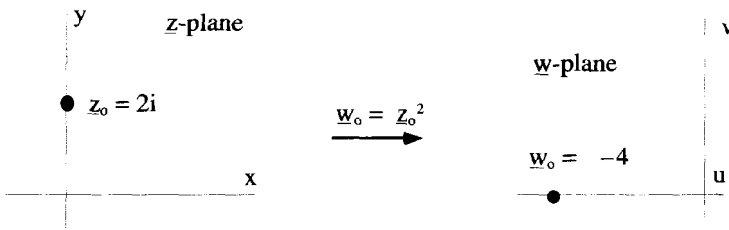
Figure 6.3 The Mapping Property of a Complex Function
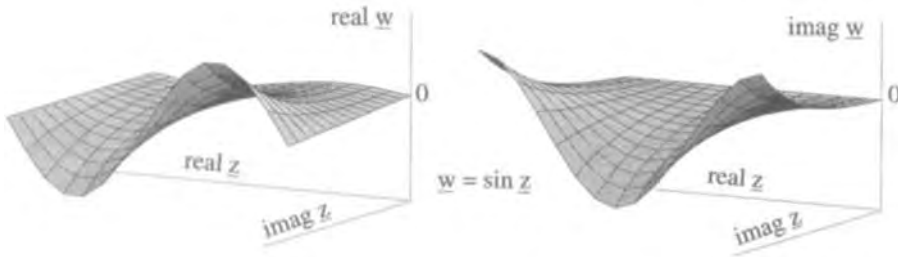
Figure 6.4 The Plots of the Real and Imaginary Parts of $\underline{w}(\underline{z})$ as Functions of x and y

6.2.2 Extending the Domain of Elementary Functions

The elementary functions of real variables can be extended to deal with complex numbers. One example of this, the complex exponential function, was already introduced in the previous section. We defined that function by extending its Taylor series expansion to cover complex quantities. Series expansions of complex functions are covered in much greater detail later in this chapter.

Trigonometric Functions We can perform a similar extension with the sine function. The Taylor series of $\sin(x)$ expanded around zero is

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \dots , \tag{6.22}$$

where x is a real variable. The *complex* sine function can be defined by extending this series to complex values,

$$\sin \underline{z} = \underline{z} - \frac{\underline{z}^3}{3!} + \frac{\underline{z}^5}{5!} - \frac{\underline{z}^7}{7!} \dots , \tag{6.23}$$

where $\underline{z}$ is now a complex quantity. This series converges for all $\underline{z}$. Similarly, the complex cosine function has the series

$$\cos \underline{z} = 1 - \frac{\underline{z}^2}{2!} + \frac{\underline{z}^4}{4!} - \frac{\underline{z}^6}{6!} + \cdots , \tag{6.24}$$

which also converges for all values of $\underline{z}$.

With these extended definitions, Euler's equation (Equation 6.8) easily generalizes to all complex quantities:

$$e^{i\underline{z}} = \cos \underline{z} + i \sin \underline{z}. \tag{6.25}$$

This allows the derivation of the commonly used expressions for the sine and cosine of complex variables:

$$\sin \underline{z} = \frac{e^{i\underline{z}} - e^{-i\underline{z}}}{2i} \tag{6.26}$$

$$\cos \underline{z} = \frac{e^{i\underline{z}} + e^{-i\underline{z}}}{2}. \tag{6.27}$$

The other trigonometric functions can be extended into the complex plane using the standard relationships. For example,

$$\tan \underline{z} = \frac{\sin \underline{z}}{\cos \underline{z}} = \frac{e^{i\underline{z}} - e^{-i\underline{z}}}{i(e^{i\underline{z}} + e^{-i\underline{z}})}. \tag{6.28}$$

Hyperbolic Functions   The hyperbolic sine and cosine functions of real variables are defined as

$$\sinh x = \frac{e^x - e^{-x}}{2} \tag{6.29}$$

$$\cosh x = \frac{e^x + e^{-x}}{2}. \tag{6.30}$$

The complex hyperbolic functions are defined by following the same rules using the complex exponential function:

$$\sinh \underline{z} = \frac{e^{\underline{z}} - e^{-\underline{z}}}{2} \tag{6.31}$$

$$\cosh \underline{z} = \frac{e^{\underline{z}} + e^{-\underline{z}}}{2}. \tag{6.32}$$

The Logarithmic Function   The definition for the logarithm of a complex variable comes directly from its polar representation:

$$\ln \underline{z} = \ln(re^{i\theta})$$
$$= \ln r + \ln(e^{i\theta})$$
$$= \ln r + i\theta. \tag{6.33}$$

6.3   DERIVATIVES OF COMPLEX FUNCTIONS

The derivative of a complex function is defined in the same way as it is for a function of a real variable:

$$\frac{d\underline{w}(\underline{z})}{d\underline{z}} \equiv \lim_{\Delta \underline{z} \to 0} \frac{\underline{w}(\underline{z} + \Delta \underline{z}) - \underline{w}(\underline{z})}{\Delta \underline{z}}. \tag{6.34}$$

This statement is not as simple as it appears, however, because $\Delta \underline{z} = \Delta x + i\Delta y$, and consequently $\Delta \underline{z}$ can be made up of any combination of Δx and $i\Delta y$. That is to say, the point $\underline{z}$ can be approached from any direction in the complex $\underline{z}$-plane. It is not obvious, and in fact not always the case, that the derivative, as defined in Equation 6.34, will have the same values for different choices of $\Delta \underline{z}$.

Functions that have a unique derivative in a finite region are said to be *analytic* in that region. As will be shown, the Cauchy Integral Theorem, the Cauchy Integral Formula, and, in fact, the whole theory of complex integration are based on the analytic properties of complex functions.

6.3.1 The Cauchy-Riemann Conditions

A necessary condition for the uniqueness of the derivative of $\underline{w}(\underline{z})$ is that we obtain the same value for it whether $\Delta\underline{z}$ is made up of purely a Δx or purely an $i\Delta y$. If $\Delta\underline{z} = \Delta x$, Equation 6.34 becomes

$$\frac{d\underline{w}(\underline{z})}{d\underline{z}} = \lim_{\Delta x \to 0} \frac{\underline{w}(x + \Delta x + iy) - \underline{w}(x + iy)}{\Delta x}. \tag{6.35}$$

On the other hand, if $\Delta\underline{z} = i\Delta y$, the derivative is

$$\frac{d\underline{w}(\underline{z})}{d\underline{z}} = \lim_{\Delta y \to 0} \frac{\underline{w}(x + iy + i\Delta y) - \underline{w}(x + iy)}{i\Delta y}. \tag{6.36}$$

If $\underline{w}(\underline{z})$ is broken up into real and imaginary parts, $\underline{w}(\underline{z}) = u(x, y) + iv(x, y)$, Equations 6.35 and 6.36 become

$$\begin{aligned}
\frac{d\underline{w}(\underline{z})}{d\underline{z}} &= \lim_{\Delta x \to 0} \frac{u(x + \Delta x, y) + iv(x + \Delta x, y) - u(x, y) - iv(x, y)}{\Delta x} \\
&= \frac{\partial u(x, y)}{\partial x} + i\frac{\partial v(x, y)}{\partial x},
\end{aligned} \tag{6.37}$$

and

$$\begin{aligned}
\frac{d\underline{w}(\underline{z})}{d\underline{z}} &= \lim_{\Delta y \to 0} \frac{u(x, y + \Delta y) + iv(x, y + \Delta y) - u(x, y) - iv(x, y)}{i\Delta y} \\
&= -i\frac{\partial u(x, y)}{\partial y} + \frac{\partial v(x, y)}{\partial y}.
\end{aligned} \tag{6.38}$$

For these derivatives to be equal, their real and imaginary parts must be equal. This requirement leads to two equations involving the derivatives of the real and imaginary parts of $\underline{w}(\underline{z})$:

$$\frac{\partial u(x, y)}{\partial x} = \frac{\partial v(x, y)}{\partial y} \tag{6.39}$$

$$\frac{\partial u(x, y)}{\partial y} = -\frac{\partial v(x, y)}{\partial x}. \tag{6.40}$$

Equations 6.39 and 6.40 are known as the Cauchy-Riemann conditions. They must be satisfied, at a point, if the derivative at that point is to be the same for $\Delta\underline{z} = \Delta x$ or $\Delta\underline{z} = i\Delta y$. Therefore, they are necessary conditions for the uniqueness of the derivative of a complex function.

6.3.2 Analytic Functions

The Cauchy-Riemann conditions are not enough to guarantee the uniqueness of the derivative. To obtain a sufficient condition, the function must meet some additional requirements.

If $\Delta \underline{z}$ is made up of an arbitrary combination of Δx and $i\Delta y$, then the derivative of $\underline{w}(\underline{z})$ can be written as:

$$\frac{d\underline{w}(\underline{z})}{d\underline{z}} = \lim_{\Delta x, \Delta y \to 0} \frac{\underline{w}(\underline{z} + \Delta x + i\Delta y) - \underline{w}(\underline{z})}{\Delta x + i\Delta y}. \tag{6.41}$$

A Taylor series expansion of the real and imaginary parts of $\underline{w}(\underline{z} + \Delta x + i\Delta y)$ about the point $z = x + iy$ gives:

$$\frac{d\underline{w}(\underline{z})}{d\underline{z}} = \lim_{\Delta x, \Delta y \to 0} \frac{(\partial u/\partial x + i\partial v/\partial x)\Delta x + (\partial u/\partial y + i\partial v/\partial y)\Delta y}{\Delta x + i\Delta y}. \tag{6.42}$$

For this equation to be valid, the partial derivatives of $u(x, y)$ and $v(x, y)$ must exist and be continuous in a *neighborhood* around the point $\underline{z}$. A neighborhood is a region of arbitrarily small, but nonzero size surrounding a point in the complex plane. Dividing the numerator and denominator of the RHS by Δx before taking the limit gives

$$\frac{d\underline{w}(\underline{z})}{d\underline{z}} = \lim_{\Delta x, \Delta y \to 0} \frac{(\partial u/\partial x + i\partial v/\partial x) + (\partial u/\partial y + i\partial v/\partial y)(\Delta y/\Delta x)}{1 + i\Delta y/\Delta x}. \tag{6.43}$$

If the derivative is to be independent of how the limit is taken with respect to Δx and Δy, the RHS of Equation 6.43 must be independent of Δx and Δy. If the Cauchy-Riemann conditions are valid not just at the point $\underline{z}$, but in the small region $\Delta x \Delta y$ about $\underline{z}$, i.e., in the neighborhood of $\underline{z}$, they can be substituted into the RHS of Equation 6.43 to give

$$\begin{aligned}
\frac{d\underline{w}(\underline{z})}{d\underline{z}} &= \lim_{\Delta x, \Delta y \to 0} \frac{(\partial u/\partial x + i\partial v/\partial x) + (-\partial v/\partial x + i\partial u/\partial x)(\Delta y/\Delta x)}{1 + i\Delta y/\Delta x} \\
&= \lim_{\Delta x, \Delta y \to 0} (\partial u/\partial x + i\partial v/\partial x) \left[\frac{1 + i\Delta y/\Delta x}{1 + i\Delta y/\Delta x} \right] \\
&= \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x},
\end{aligned} \tag{6.44}$$

which is independent of $\Delta \underline{z}$.

Consequently, if the partial derivatives of $u(x, y)$ and $v(x, y)$ exist, are continuous, and obey the Cauchy-Riemann conditions in a neighborhood around a point $\underline{z}$, the derivative of the function at that point is guaranteed to be unique. Functions that obey all these conditions are called analytic. A function may be analytic in the entire $\underline{z}$-plane or in isolated regions. But because the Cauchy-Riemann conditions must be satisfied in a neighborhood of the complex plane, a function cannot be analytic at an isolated point or along a line. The sum, product, quotient (as long as the denominator is not zero) and composition of two functions that are analytic in a region are analytic in that same region. Proofs of these statements are left as an exercise for the reader.

There are many useful complex variable techniques which require the use of analytic functions. In fact, the rest of the material in this chapter concerns the properties of analytic functions.

Example 6.1 First consider the function $\underline{w} = \underline{z}^2$ with

$$u(x, y) = x^2 - y^2 \tag{6.45}$$

$$v(x, y) = 2xy. \tag{6.46}$$

Taking the partial derivatives,

$$\frac{\partial u(x, y)}{\partial x} = 2x = \frac{\partial v(x, y)}{\partial y} \tag{6.47}$$

and

$$\frac{\partial u(x, y)}{\partial y} = -2y = -\frac{\partial v(x, y)}{\partial x}. \tag{6.48}$$

This function satisfies the Cauchy-Riemann conditions everywhere in the finite complex $\underline{z}$-plane. From either Equation 6.37 or 6.38,

$$\frac{d\underline{w}}{d\underline{z}} = \frac{\partial u(x, y)}{\partial x} + i\frac{\partial v(x, y)}{\partial x} = \frac{\partial v(x, y)}{\partial y} - i\frac{\partial u(x, y)}{\partial y}$$

$$= 2x + i2y \tag{6.49}$$

$$= 2\underline{z}.$$

Notice that this derivative could have been obtained by treating $\underline{z}$ in the same manner as a real, differential variable, i.e.,

$$\frac{d\underline{w}}{d\underline{z}} = \frac{d(\underline{z}^2)}{d\underline{z}} = 2\underline{z}. \tag{6.50}$$

Example 6.2 As another example, consider the function $\underline{w} = 1/\underline{z}$. In this case, the real and imaginary parts of $\underline{w}$ are obtained by using $\underline{z}^*$,

$$\underline{w} = \frac{1}{\underline{z}} = \frac{1}{\underline{z}} \cdot \frac{\underline{z}^*}{\underline{z}^*} = \frac{x - iy}{x^2 + y^2}, \tag{6.51}$$

so that

$$u(x, y) = \frac{x}{x^2 + y^2} \tag{6.52}$$

$$v(x, y) = \frac{-y}{x^2 + y^2}. \tag{6.53}$$

A check of the partial derivatives shows that the Cauchy-Riemann conditions are satisfied everywhere, except at $\underline{z} = 0$, where the partial derivatives blow up. Therefore, $1/\underline{z}$ is analytic everywhere except at $\underline{z} = 0$, and its derivative is

$$\frac{d(1/\underline{z})}{d\underline{z}} = -\frac{1}{\underline{z}^2}. \tag{6.54}$$

Example 6.3   As a final example, consider the function $\underline{w} = \underline{z}\,\underline{z}^*$, so that

$$\underline{w} = (x + iy)(x - iy) = x^2 + y^2. \tag{6.55}$$

The real and imaginary parts are

$$u(x, y) = x^2 + y^2 \tag{6.56}$$

$$v(x, y) = 0 \tag{6.57}$$

so that

$$\frac{\partial u(x, y)}{\partial x} = 2x \neq \frac{\partial v(x, y)}{\partial y} = 0 \tag{6.58}$$

and

$$\frac{\partial u(x, y)}{\partial y} = 2y \neq -\frac{\partial v(x, y)}{\partial x} = 0. \tag{6.59}$$

The Cauchy-Riemann conditions are *not* satisfied, except at the origin. Because this a single, isolated point, the function $\underline{w} = \underline{z}\,\underline{z}^*$ is not an analytic function anywhere. This is made more obvious when the derivative of this function is formed. Using Equation 6.37, the derivative is

$$\frac{d(\underline{z}\,\underline{z}^*)}{d\underline{z}} = 2x = \underline{z} + \underline{z}^*, \tag{6.60}$$

while using Equation 6.38,

$$\frac{d(\underline{z}\,\underline{z}^*)}{d\underline{z}} = -i2y = \underline{z}^* - \underline{z}. \tag{6.61}$$

6.3.3   Derivative Formulae

The rules you have learned for real derivatives hold for the derivatives of complex analytic functions. The standard chain, product, and quotient rules still apply. All the elementary functions have the same derivative formulae. For example:

$$\frac{d}{d\underline{z}}\left[e^{i\cos(\underline{z})}\right] = ie^{i\cos(\underline{z})}\frac{d}{d\underline{z}}[\cos(\underline{z})] = -ie^{i\cos(\underline{z})}\sin\underline{z}. \tag{6.62}$$

6.4   THE CAUCHY INTEGRAL THEOREM

An important consequence of using analytic functions is the Cauchy Integral Theorem. Let $\underline{w}(\underline{z})$ be a complex function which is analytic in some region of the complex
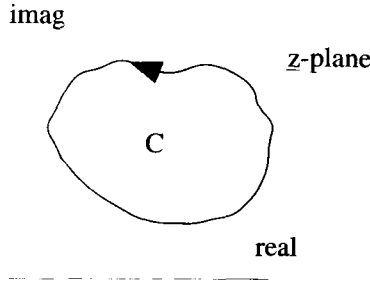
imag

z-plane

C

real

Figure 6.5 The Closed Cauchy Integral Path

plane. Consider an integral

$$\oint_C d\underline{z}\, w(\underline{z}) \tag{6.63}$$

along a closed contour C which lies entirely within the analytic region, as shown in Figure 6.5. Because $\underline{w}(\underline{z}) = u(x, y) + iv(x, y)$ and $d\underline{z} = dx + idy$, Equation 6.63 can be rewritten as

$$\oint_C d\underline{z}\, \underline{w}(\underline{z}) = \oint_C \left(dx + idy \right) [u(x, y) + iv(x, y)] \tag{6.64}$$

$$= \oint_C [dx\, u(x, y) - dy\, v(x, y)] + i \oint_C [dx\, v(x, y) + dy\, u(x, y)].$$

Stokes's theorem

$$\oint_C d\overline{\mathbf{r}} \cdot \overline{\mathbf{V}} = \int_S d\overline{\boldsymbol{\sigma}} \cdot \overline{\mathbf{V}} \times \overline{\mathbf{V}} \tag{6.65}$$

can now be applied to both integrals on the RHS of Equation 6.64. For a closed line integral confined to the xy-plane, Equation 6.65 becomes

$$\oint_C dx\, V_x + dy\, V_y = \int_S dx\, dy \left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} \right). \tag{6.66}$$

The first integral on the RHS of Equation 6.64 is handled by identifying $V_x = u(x, y)$, $V_y = -v(x, y)$, and S as the surface enclosed by C, so

$$\oint_C [dx\, u(x, y) - dy\, v(x, y)] = -\int_S dx\, dy \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right). \tag{6.67}$$

The second integral of Equation 6.64 is handled in a similar manner, by identifying $V_x = v(x, y)$ and $V_y = u(x, y)$, so

$$\oint_C [dx\, v(x, y) + dy\, u(x, y)] = \int_S dx\, dy \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \right). \tag{6.68}$$

Because $\underline{w}(\underline{z})$ is analytic and obeys the Cauchy-Riemann conditions inside the entire closed contour, $\partial u/\partial x = \partial v/\partial y$ and $\partial u/\partial y = -\partial v/\partial x$, and we are left with the elegant result that

$$\oint_C d\underline{z}\, \underline{w}(\underline{z}) = 0. \tag{6.69}$$

6.5   CONTOUR DEFORMATION

There is an immediate, practical consequence of the Cauchy Integral Theorem. The contour of an complex integral can be arbitrarily deformed through an analytic region without changing the value of the integral.

Consider the integral

$$I = \int_{C_o} d\underline{z}\, \underline{w}(\underline{z}) \tag{6.70}$$

along the contour C_o from a to b, drawn as the solid line in Figure 6.6(a). If we add to this an integral with the same integrand, but around the closed contour C, shown as the dotted line in Figure 6.6(a), the result will still be I, as long as the integrand is analytic in the entire region inside C:

$$I = \int_{C_o} d\underline{z}\, \underline{w}(\underline{z}) + \oint_C d\underline{z}\, \underline{w}(\underline{z}). \tag{6.71}$$
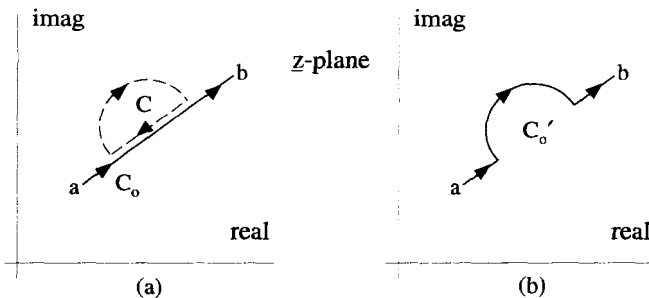

Figure 6.6   Contour Deformation

But notice the part of C that lies along the original contour has an equal, but opposite contribution to the value, and the contributions from the two segments cancel, leaving the contour shown in Figure 6.6(b). Therefore,

$$I = \int_{C_o} dz\, \underline{w}(\underline{z}) = \int_{C_o'} dz\, \underline{w}(\underline{z}). \tag{6.72}$$

Deformations of this kind will be very handy in the discussions that follow.

6.6   THE CAUCHY INTEGRAL FORMULA

The Cauchy Integral Formula is a natural extension of the Cauchy Integral Theorem discussed previously. Consider the integral

$$I_1 = \oint_{C_1} dz \frac{f(z)}{\underline{z} - \underline{z}_o}, \tag{6.73}$$

where $f(\underline{z})$ is analytic everywhere in the $\underline{z}$-plane, and C_1 is a closed contour that does not include the point $\underline{z}_o$, as depicted in Figure 6.7. It has already been shown that the function $1/\underline{z}$ is analytic everywhere except at $\underline{z} = 0$. Therefore the function $1/(\underline{z} - \underline{z}_o)$ is analytic everywhere, except at $\underline{z} = \underline{z}_o$. Because C_1 does not enclose the point $\underline{z} = \underline{z}_o$, the integrand in Equation 6.73 is analytic everywhere inside C_1, and by Cauchy's Integral Theorem,

$$I_1 = \oint_{C_1} dz \frac{f(z)}{\underline{z} - \underline{z}_o} = 0. \tag{6.74}$$

Now consider a second integral,

$$I_2 = \oint_{C_2} dz \frac{f(z)}{\underline{z} - \underline{z}_o}, \tag{6.75}$$

similar to the first, except now the contour C_2 encloses $\underline{z}_o$, as shown in Figure 6.8. The integrand in Equation 6.75 is not analytic inside C_2, so we cannot invoke the Cauchy
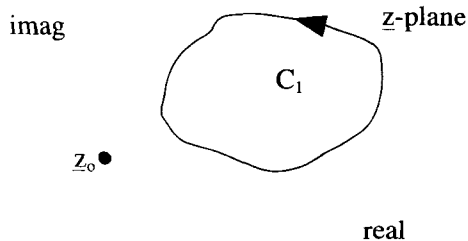

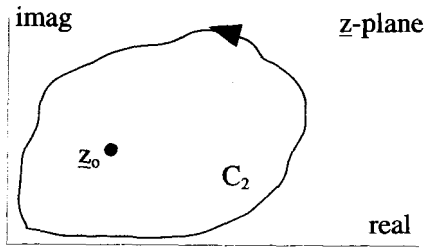**Figure 6.7**   The Contour C_1 for the Cauchy Integral Formula

Figure 6.8 The Contour C_2 for the Cauchy Integral Formula

Integral Theorem to argue that $I_2 = 0$. However, the integrand is analytic everywhere, except at the point $\underline{z} = \underline{z}_o$, so we can deform the contour into a infinitesimal circle of radius ϵ, centered at $\underline{z}_o$, without affecting the value:

$$I_2 = \lim_{\epsilon \to 0} \oint_{C_o} d\underline{z} \frac{f(\underline{z})}{\underline{z} - \underline{z}_o}. \tag{6.76}$$

This deformation is depicted in Figure 6.9.

The integral expression of Equation 6.76 can be visualized most easily with the use of phasors, which are essentially vectors in the complex plane. Consider Figure 6.10. In Equation 6.76, $\underline{z}$ lies on the contour C_o and is represented by a phasor from the origin to that point. The point $\underline{z}_o$ is represented by another phasor from the origin to $\underline{z}_o$. In the same way, the quantity $\underline{z} - \underline{z}_o$ is represented by a phasor that goes from $\underline{z}_o$ to $\underline{z}$. If the angle ϕ is defined as shown, using polar notation we have

$$\underline{z} - \underline{z}_o = \epsilon e^{i\phi}. \tag{6.77}$$

On C_o, ϵ is a constant so

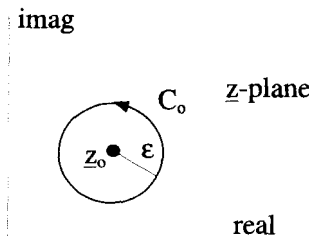$$d\underline{z} = i\epsilon e^{i\phi} d\phi, \tag{6.78}$$

Figure 6.9 The Equivalent Circular Contour for the Cauchy Integral Formula
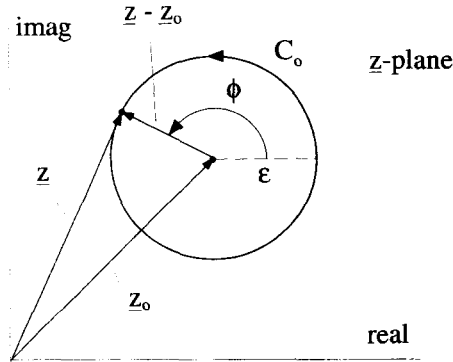
Figure 6.10   Phasors for the Cauchy Formula Integration

and ϕ goes from 0 to 2π. The integral in Equation 6.76 becomes

$$I_2 = \lim_{\epsilon \to 0} \int_0^{2\pi} d\phi \, \frac{f(\underline{z}_o + \epsilon e^{i\phi})}{\epsilon e^{i\phi}} \, i\epsilon e^{i\theta}. \tag{6.79}$$

As $\epsilon \to 0$, $\underline{f}(\underline{z}_o + \epsilon e^{i\phi})$ goes to $\underline{f}(\underline{z}_o)$ and can be taken outside the integral:

$$I_2 = i \, \underline{f}(\underline{z}_o) \int_0^{2\pi} d\phi = 2\pi i \, \underline{f}(\underline{z}_o). \tag{6.80}$$

Our final result is

$$\oint_C \frac{f(\underline{z})}{\underline{z} - \underline{z}_o} d\underline{z} = \begin{cases} 2\pi i \, \underline{f}(\underline{z}_o) & \underline{z}_o \text{ inside } C \\ 0 & \underline{z}_o \text{ outside } C \end{cases}, \tag{6.81}$$

where C is any closed, counterclockwise path that encloses $\underline{z}_o$, and $\underline{f}(\underline{z})$ is analytic inside C.

Equation 6.81 is the simplest form of Cauchy's Integral Formula. A more general result can be obtained by considering the point $\underline{z}_o$ to be a variable. If we restrict $\underline{z}_o$ to always lie inside the contour C, then Equation 6.81 can be differentiated with respect to $\underline{z}_o$ to give

$$\frac{d \, \underline{f}(\underline{z}_o)}{d \underline{z}_o} = \frac{1}{2\pi i} \oint_C d\underline{z} \frac{f(\underline{z})}{(\underline{z} - \underline{z}_o)^2}. \tag{6.82}$$

In fact, Equation 6.81 can be differentiated any number of times, to give the most general form of Cauchy's Integral Formula:

$$\frac{d^n \, \underline{f}(\underline{z}_o)}{d \underline{z}_o^n} = \frac{n!}{2\pi i} \oint_C d\underline{z} \frac{f(\underline{z})}{(\underline{z} - \underline{z}_o)^{n+1}}. \tag{6.83}$$

Here again, C is any closed, counterclockwise contour, $\underline{z}_o$ is any point inside C, and $\underline{f}(\underline{z})$ is analytic everywhere inside C.

Equation 6.83 says something quite important about the nature of analytic functions. If the value of an analytic function is known everywhere on some closed contour C, the value of that function and all its derivatives are uniquely specified everywhere inside C.

6.7   TAYLOR AND LAURENT SERIES

Just as with real functions, complex functions can be expanded using Taylor series. An extension of the standard Taylor series, called the Laurent series, is also frequently used to expand complex functions. Series expansions play a crucial role in the theory of complex functions, because they provide the fundamental tools for deriving and using the theory of residues, described in the next section.

6.7.1   Convergence

Before discussing Taylor and Laurent expansions for complex functions, it is appropriate to say a few words about the convergence of complex series. We assume you are already familiar with the convergence properties of real series, so we will not attempt to cover that topic with any sort of mathematical rigor.

An infinite summation of real numbers x_n is said to converge to the value S if

$$S = \sum_{n=0}^{\infty} x_n \rightarrow \text{A Finite Real Number}. \tag{6.84}$$

The series in Equation 6.84, is said to converge absolutely if

$$\sum_{n=0}^{\infty} |x_n| \rightarrow \text{A Finite Real Number}. \tag{6.85}$$

Because

$$\sum_{n=0}^{\infty} |x_n| \geq \sum_{n=0}^{\infty} x_n, \tag{6.86}$$

absolute convergence is a more rigid requirement than convergence. That is, if a series converges absolutely it will always converge. However, a series may converge even though it does not converge absolutely. For example, the alternating harmonic series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots \tag{6.87}$$

converges to $\ln 2$, but the harmonic series

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \cdots \tag{6.88}$$

diverges.

The ratio test can be used to establish the absolute convergence of a series. That is, if the terms of an infinite series satisfy

$$\lim_{n \to \infty} \left| \frac{x_{n+1}}{x_n} \right| < 1, \tag{6.89}$$

the series converges absolutely. Again, a series may converge even though it does not satisfy the ratio test. There are plenty of other tests for convergence, but the ratio test is the only one we need for our discussions.

Things are not too different when we consider infinite series of complex numbers. A summation of complex numbers $\underline{z}_n$ converges to the complex value $\underline{S}$ if

$$\underline{S} = \sum_{n=0}^{\infty} \underline{z}_n \to \text{A Finite Complex Number.} \tag{6.90}$$

For convergence of the infinite complex series, both the real and imaginary parts must converge. That is, if we write $\underline{z}_n = x_n + iy_n$ and $\underline{S} = S_r + iS_i$, we must have

$$S_r = \sum_{n=0}^{\infty} x_n \to \text{A Finite Real Number}$$
$$S_i = \sum_{n=0}^{\infty} y_n \to \text{A Finite Real Number.} \tag{6.91}$$

The complex series defined in Equation 6.90 converges absolutely if

$$\sum_{n=0}^{\infty} \left| \underline{z}_n \right| \to \text{A Finite Complex Number.} \tag{6.92}$$

Because

$$\sum_{n=0}^{\infty} \left| \underline{z}_n \right| \ge \sum_{n=0}^{\infty} x_n \tag{6.93}$$

and

$$\sum_{n=0}^{\infty} \left| \underline{z}_n \right| \ge \sum_{n=0}^{\infty} y_n, \tag{6.94}$$

if the series converges absolutely, the two series in Equations 6.91 will both converge, and consequently the original complex series in Equation 6.90 must also converge. The absolute convergence of a complex series can also be established by the ratio test. That is, if the terms of the series satisfy

$$\lim_{n \to \infty} \left| \frac{\underline{z}_{n+1}}{\underline{z}_n} \right| < 1, \tag{6.95}$$

the series will converge absolutely.

6.7.2 Taylor and Laurent Series of Real Functions

Before dealing with expansions for complex functions, Taylor and Laurent series for real functions will be reviewed. Recall that the Taylor series expansion of $f(x)$, a real function of the variable x, around the point x_o is

$$f(x) = \sum_{n=0}^{\infty} c_n(x - x_o)^n = c_o + c_1(x - x_o) + c_2(x - x_o)^2 + \cdots. \qquad (6.96)$$

The coefficients c_n can be obtained from the expression

$$c_n = \frac{1}{n!}\frac{d^n f(x)}{dx^n}\bigg|_{x=x_o} \qquad (6.97)$$

and depend upon both the function $f(x)$ and the expansion point x_o. The region of convergence of this series can be determined by using the ratio test. The series on the RHS of Equation 6.96 converges if

$$\lim_{n\to\infty}\left|\frac{c_n(x - x_o)^n}{c_{n-1}(x - x_o)^{n-1}}\right| < 1. \qquad (6.98)$$

In general a Taylor series will converge for values of x in a symmetric range around x_o, given by $|x - x_o| < \Delta$, where Δ depends upon both the expansion point and the function $f(x)$.

A Laurent series expansion differs from a Taylor series because of the additional terms with negative n values. An expansion of $f(x)$ around x_o using a Laurent series has the form

$$f(x) = \sum_{n=-\infty}^{+\infty} c_n(x - x_o)^n \qquad (6.99)$$

$$= \cdots + \frac{c_{-2}}{(x - x_o)^2} + \frac{c_{-1}}{(x - x_o)} + c_o + c_1(x - x_o) + c_2(x - x_o)^2 + \cdots.$$

Convergence for a Laurent series is slightly more complicated to verify than for a Taylor series, because we must check the convergence of both the positive and negative terms. In many cases, the negative terms of the series terminate at some finite value of n, and we only need to worry about the convergence of the positive terms. For the general case, however, we can perform two ratio tests, one for the positive terms and one for the negative terms. Clearly, if there are any terms with negative n values, the series cannot converge at $x = x_o$. In general, a Laurent series will converge in a symmetric range of x given by $\Delta_{min} < |x - x_o| < \Delta_{max}$, where Δ_{min} and Δ_{max} are the minimum and maximum allowed distances from the expansion point.

$$\begin{array}{r} 1 + \underline{z} + \underline{z}^2 + \quad \cdots \quad + \underline{z}^m \\ \hline 1 - \underline{z} \, \big|\, 1 \qquad\qquad\qquad\qquad\qquad - \underline{z}^{m+1} \\ \underline{1 - \underline{z}} \\ \underline{z} \\ \underline{1 - \underline{z}^2} \\ \underline{z}^2 \\ \underline{z}^2 - \underline{z}^3 \\ \ddots \\ \underline{z}^m - \underline{z}^{m+1} \\ \underline{\underline{z}^m - \underline{z}^{m+1}} \\ 0 \end{array}$$

Figure 6.11 Polynomial Long Division

6.7.3 A Basic Complex Taylor Expansion

As part of the derivation of complex Taylor and Laurent series, the following basic complex Taylor expansion will prove to be valuable. This expansion,

$$\frac{1}{1 - \underline{z}} = 1 + \underline{z} + \underline{z}^2 + \underline{z}^3 + \cdots = \sum_{n=0}^{\infty} \underline{z}^n, \tag{6.100}$$

converges absolutely for all $|\underline{z}| < 1$, a fact easily shown by applying Equation 6.95, the ratio test for complex series. This result derives from the $m \to \infty$ limit of the identity

$$1 + \underline{z} + \underline{z}^2 + \underline{z}^3 + \cdots + \underline{z}^m = \frac{1 - \underline{z}^{m+1}}{1 - \underline{z}}, \tag{6.101}$$

which can be proved by multiplying both sides of this expression by $(1 - \underline{z})$, or by the long division shown in Figure 6.11.

6.8 THE COMPLEX TAYLOR SERIES

Similar to a real Taylor series, a complex Taylor series attempts to expand a complex function around a point, using a sum of complex terms:

$$\underline{f}(\underline{z}) = \sum_{n=0}^{\infty} \underline{c}_n (\underline{z} - \underline{z}_o)^n = \underline{c}_o + \underline{c}_1 (\underline{z} - \underline{z}_o) + \underline{c}_2 (\underline{z} - \underline{z}_o)^2 + \cdots. \tag{6.102}$$

The $\underline{c}_n$ coefficients and the region of convergence of the series depend on the function $\underline{f}(\underline{z})$ and the expansion point $\underline{z}_o$. We will show how these coefficients and the region of convergence can be obtained in two ways, first by using Cauchy integra-

tion techniques, and then by applying the basic complex Taylor series developed in Section 6.7.3.

6.8.1   The Taylor Coefficients from Cauchy Integration

The process of determining the coefficients from a Cauchy integration begins by selecting an expansion point z_o and a counterclockwise, circular contour C centered around the expansion point, as shown in Figure 6.12. The function $f(z)$ and its analytic properties are presumed to be known. Assuming $f(z)$ is analytic everywhere inside C, Cauchy's Integral Formula (Equation 6.81) states that

$$f(z) = \frac{1}{2\pi i} \oint_C dz' \frac{f(z')}{z' - z} \quad , \tag{6.103}$$

where z is any point inside C, and z' is the complex variable of integration that lies along C. z_o can be inserted into the denominator of the integrand and manipulated as follows:

$$f(z) = \frac{1}{2\pi i} \oint_C dz' \frac{f(z')}{(z' - z_o) - (z - z_o)}$$

$$= \frac{1}{2\pi i} \oint_C dz' \left[\frac{f(z')}{z' - z_o} \left(\frac{1}{1 - \frac{z - z_o}{z' - z_o}} \right) \right]. \tag{6.104}$$

Because $|z - z_o| < |z' - z_o|$ for all z inside C, the basic complex Taylor series of Equation 6.100 implies that

$$\frac{1}{1 - \frac{z - z_o}{z' - z_o}} = \sum_{n=0}^{\infty} \left(\frac{z - z_o}{z' - z_o} \right)^n, \tag{6.105}$$
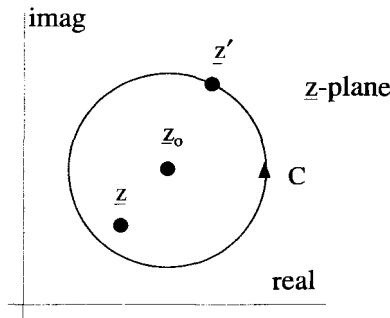

Figure 6.12   Contour for Taylor Series Coefficients

which converges for all $\underline{z}$ inside C. Substitution of Equation 6.105 into 6.104 gives

$$\underline{f}(\underline{z}) = \frac{1}{2\pi i} \oint_C d\underline{z}' \left[\frac{f(\underline{z}')}{\underline{z}' - \underline{z}_o} \sum_{n=0}^{\infty} \left(\frac{\underline{z} - \underline{z}_o}{\underline{z}' - \underline{z}_o} \right)^n \right]. \qquad (6.106)$$

We will now interchange the order of integration and summation in Equation 6.106. It is not obvious that this is mathematically allowed, because we are manipulating an infinite series. An important theorem, however, says that the integration and summation of a series, or for that matter the differentiation and summation of a series, can be interchanged if the series in question converges uniformly. Uniform convergence is a property of infinite series of functions of one or more independent variables. We will not cover this topic in this text, but we refer you to almost any book on advanced calculus. See, for instance, Kaplan, *Advanced Calculus* or Sokolnikoff and Redheffer, *Mathematics of Physics and Modern Engineering*. The interchange of the integration and summation gives

$$\underline{f}(\underline{z}) = \sum_{n=0}^{\infty} \left[\frac{1}{2\pi i} \oint_C d\underline{z}' \frac{f(\underline{z}')}{(\underline{z}' - \underline{z}_o)^{n+1}} \right] (\underline{z} - \underline{z}_o)^n. \qquad (6.107)$$

Comparing Equation 6.107 with Equation 6.102, the general form of the Taylor series for $\underline{f}(\underline{z})$, identifies the coefficients as

$$\underline{c}_n = \frac{1}{2\pi i} \oint_C d\underline{z}' \frac{f(\underline{z}')}{(\underline{z}' - \underline{z}_o)^{n+1}}. \qquad (6.108)$$

Using Equation 6.83, the general form of Cauchy's Integral Formula, the coefficients can be recognized as

$$\underline{c}_n = \frac{1}{n!} \frac{d^n \underline{f}(\underline{z})}{d\underline{z}^n} \bigg|_{\underline{z}=\underline{z}_o}, \qquad (6.109)$$

which is a satisfying extension of Equation 6.97 for the coefficients of real Taylor series. These coefficients will generate a complex Taylor series that will converge to $\underline{f}(\underline{z})$ for all $\underline{z}$ inside C. The convergence properties of the series are best demonstrated by the example that follows.

Example 6.4   As an example of this process, let's obtain the coefficients of the Taylor series for the complex function

$$\underline{f}(\underline{z}) = \frac{1}{\underline{z} - \underline{a}}. \qquad (6.110)$$

This function is analytic everywhere in the complex plane, except at $\underline{z} = \underline{a}$. The Taylor series is to be expanded about $\underline{z} = \underline{z}_o$. These points are shown in Figure 6.13. The circular path for the Cauchy integration is also shown in this figure. It is centered at $\underline{z}_o$ and has a radius such that $1/(\underline{z} - \underline{a})$ is analytic everywhere inside. In other words, if $\underline{z}'$ is a point along the contour, $|\underline{z}' - \underline{z}_o| < |\underline{a} - \underline{z}_o|$. The coefficients for the
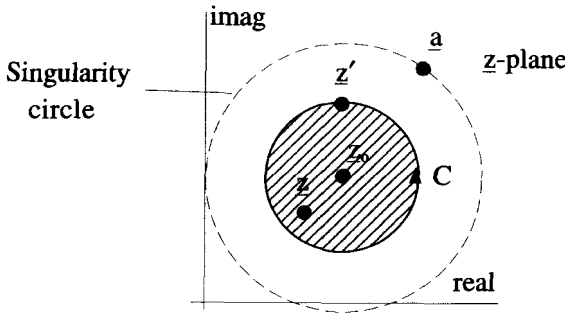
Figure 6.13   The Complex Plane for Determining the Taylor Coefficients for $1/(\underline{z} - \underline{a})$

Taylor series can now be obtained from Equation 6.109 as

$$
\begin{aligned}
\underline{c}_n &= \frac{1}{n!} \left. \frac{d^n \underline{f}(\underline{z})}{d\underline{z}^n} \right|_{\underline{z}=\underline{z}_o} \\
&= -\left(\frac{1}{\underline{a} - \underline{z}_o} \right)^{n+1}
\end{aligned}
\tag{6.111}
$$

so that

$$
\frac{1}{\underline{z} - \underline{a}} = \sum_{n=0}^{\infty} - \left(\frac{1}{\underline{a} - \underline{z}_o} \right)^{n+1} (\underline{z} - \underline{z}_o)^n .
\tag{6.112}
$$

This series converges in the crosshatched region of Figure 6.13 where $|\underline{z} - \underline{z}_o| < |\underline{z}' - \underline{z}_o|$. In that same figure, we have drawn a dashed circle, centered at $\underline{z}_0$ and passing through the singularity at $\underline{a}$. This circle, which we call a *singularity circle*, is the largest radius that the contour C can have, while still keeping $\underline{f}(\underline{z})$ analytic everywhere inside. Consequently, the region of convergence for the series in Equation 6.111 can be made to extend up to the singularity circle, as shown in Figure 6.14. If a value of $\underline{z}$ is chosen
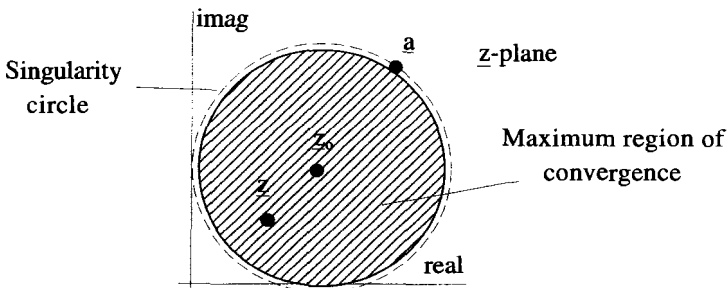

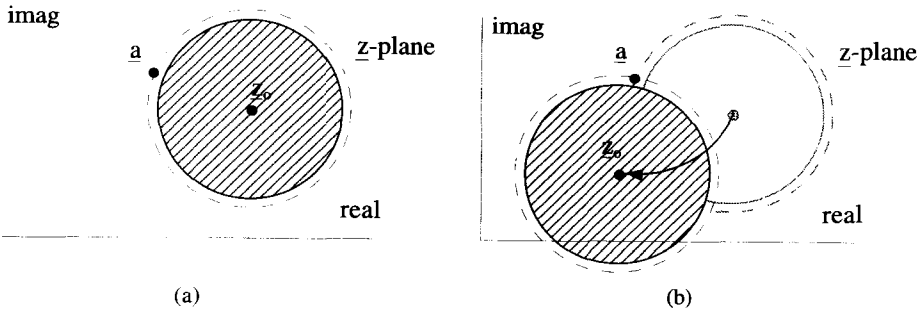**Figure 6.14**   Region of Convergence

Figure 6.15 The z_o Dependence of the Convergence Region

that lies inside this singularity circle, and is inserted into the LHS of Equation 6.112, we will obtain the same finite value that we get from inserting this same z into the RHS. On the other hand, if we select a point outside of the singularity circle, and use it in the same equation, the LHS will give a finite number, while the RHS will diverge.

If we have a more complicated function, with more than one nonanalytic point, the region of convergence is inside the smallest of all the singularity circles. Also, notice that the Taylor coefficients and the region of convergence depend on the choice of z_o. By selecting different z_o's, different Taylor series can be generated. Figure 6.15 shows two different expansion points and their corresponding regions of convergence. In this way, a Taylor series can be generated for this function that is valid for any z, except $z = a$. From this picture it can also be seen that no single Taylor series can be made to converge in a region that surrounds a singularity point. This will be an important point in our discussion of Laurent series.

6.8.2 The Taylor Coefficients Using the Basic Taylor Series

Many times it is possible to determine the Taylor series coefficients and the region of convergence without using the technique described above, but instead by manipulating $f(z)$ to be in the form of the basic Taylor series given in Equation 6.100. This is best demonstrated by the following examples.

Example 6.5 Again consider the function

$$f(z) = \frac{1}{z - a},\tag{6.113}$$

where a is a complex constant. Dividing the numerator and denominator by $-a$ gives

$$f(z) = \frac{-1/a}{1 - z/a}.\tag{6.114}$$

As long as $|\underline{z}/\underline{a}| < 1$, the denominator can be expanded using Equation 6.100 to give

$$\frac{1}{\underline{z} - \underline{a}} = \frac{-1}{\underline{a}} \left[1 + \frac{\underline{z}}{\underline{a}} + \left(\frac{\underline{z}}{\underline{a}} \right)^2 + \left(\frac{\underline{z}}{\underline{a}} \right)^3 + \cdots \right]. \tag{6.115}$$

But Equation 6.115 is already in the form of a complex Taylor series, with an expansion point around the origin:

$$\frac{1}{\underline{z} - \underline{a}} = \sum_{n=0}^{\infty} \underline{c}_n \underline{z}^n \tag{6.116}$$

with

$$\underline{c}_n = -\left(\frac{1}{\underline{a}} \right)^{n+1}. \tag{6.117}$$

Convergence of this series can be checked by a ratio test

$$\lim_{n \to \infty} \left| \frac{\underline{c}_{n+1} \underline{z}^{n+1}}{\underline{c}_n \underline{z}^n} \right| < 1, \tag{6.118}$$

which gives the condition

$$|\underline{z}| < |\underline{a}|, \tag{6.119}$$

which you might have guessed from the condition on the series expansion used in Equation 6.115.

Notice the convergence region is a circle centered at the expansion point with a radius of $|\underline{a}|$. This confirms the previous discussion which led to Figure 6.14.

Example 6.6 Now let's seek a Taylor series for the same function, but this time expanded about an arbitrary point $\underline{z} = \underline{z}_o$. The series must take the form

$$\frac{1}{\underline{z} - \underline{a}} = \sum_{n=0}^{+\infty} \underline{c}_n (\underline{z} - \underline{z}_o)^n. \tag{6.120}$$

If we manipulate the function as follows:

$$\frac{1}{\underline{z} - \underline{a}} = \frac{1}{(\underline{z} - \underline{z}_o) - (\underline{a} - \underline{z}_o)} = \left(\frac{-1}{\underline{a} - \underline{z}_o} \right) \left(\frac{1}{1 - \frac{\underline{z} - \underline{z}_o}{\underline{a} - \underline{z}_o}} \right), \tag{6.121}$$

we can use the same technique as before, and write

$$\frac{1}{\underline{z} - \underline{a}} = \left(\frac{-1}{\underline{a} - \underline{z}_o} \right) \left[1 + \frac{\underline{z} - \underline{z}_o}{\underline{a} - \underline{z}_o} + \frac{(\underline{z} - \underline{z}_o)^2}{(\underline{a} - \underline{z}_o)^2} + \cdots \right], \tag{6.122}$$

with the region of convergence

$$\left| \frac{\underline{z} - \underline{z}_o}{\underline{a} - \underline{z}_o} \right| < 1, \tag{6.123}$$

which is the same region depicted in Figure 6.14. Thus we find the coefficients for this expansion are

$$c_n = -\left(\frac{1}{a - z_o}\right)^{n+1}. \tag{6.124}$$

Once again, the series converges in a circular region centered at z_o. The convergence radius is the distance from the expansion point to the nonanalytic point a. Notice that these are the same coefficients and region of convergence we obtained using the Cauchy integration approach.

6.9 THE COMPLEX LAURENT SERIES

A complex Laurent series has the form

$$f(z) = \sum_{n=-\infty}^{\infty} c_n (z - z_o)^n, \tag{6.125}$$

and is distinguished from a Taylor series because it contains negative n terms. As with the Taylor series, the coefficients and the region of convergence depend upon the function $f(z)$ and the expansion point z_o. These coefficients and the region of convergence for the Laurent series will be obtained from a Cauchy integral approach, and also by using the basic complex Taylor series developed in Equation 6.100. Laurent series are useful when an expansion is desired that converges in a region surrounding a nonanalytic point of $f(z)$.

6.9.1 The Laurent Coefficients from Cauchy Integration

Imagine we wanted to generate a series expansion for the function $1/(z - a)$ about z_o, as we did in the previous section, but this time, we want an expansion which is valid in a region outside of the singularity circle of a. We cannot use the contour shown in Figure 6.16, or any such contour which surrounds the singularity, and expect to

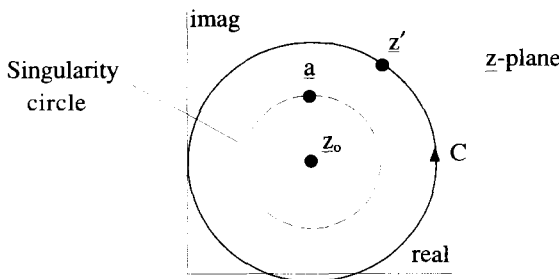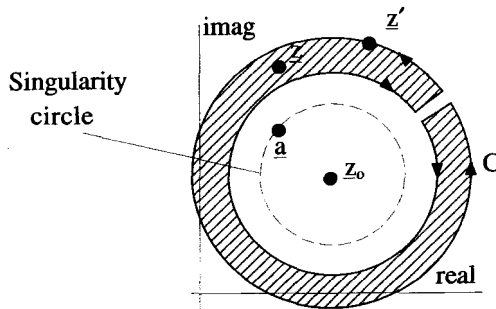**Figure 6.16** Invalid Contour for Expanding for $1/(z - a)$ around z_o

Figure 6.17   Contour for Calculating Laurent Coefficients

generate a Taylor series using Equation 6.108, because that equation requires that $\underline{f}(\underline{z})$ be analytic at all points inside C.

Instead, consider the contour shown in Figure 6.17. In the crosshatched region interior to C, the function $\underline{f}(\underline{z})$ is analytic. Therefore, according to Cauchy's Integral Formula,

$$\underline{f}(\underline{z}) = \frac{1}{2\pi i} \oint_C d\underline{z}' \frac{\underline{f}(\underline{z}')}{\underline{z}' - \underline{z}}. \tag{6.126}$$

Using contour deformation, the two straight-end pieces of C can be brought arbitrarily close together, so their contributions to the integral cancel. Therefore, we can rewrite this integral as

$$\underline{f}(\underline{z}) = \frac{1}{2\pi i} \oint_{C_1} d\underline{z}' \frac{\underline{f}(\underline{z}')}{\underline{z}' - \underline{z}} - \frac{1}{2\pi i} \oint_{C_2} d\underline{z}' \frac{\underline{f}(\underline{z}')}{\underline{z}' - \underline{z}}, \tag{6.127}$$

where C_1 and C_2 are shown in Figure 6.18. Notice the sneaky introduction of a minus sign into the contour C_2, so both contours are now counterclockwise. Again, we require C_1 and C_2 to be circular, so we can use a similar expansion to the one used with the Taylor series.
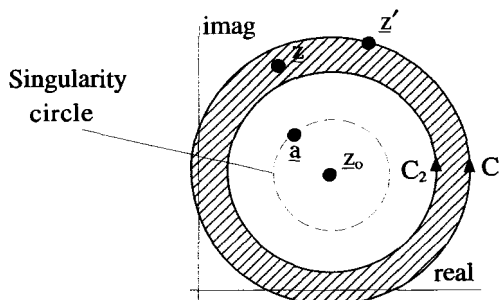

Figure 6.18   Double Circular Contour for Laurent Coefficient Integration

Next manipulate the integrals of Equation 6.127. First insert $\underline{z}_o$,

$$\underline{f}(\underline{z}) = \frac{1}{2\pi i} \oint_{C_1} d\underline{z}' \frac{\underline{f}(\underline{z}')}{(\underline{z}' - \underline{z}_o) - (\underline{z} - \underline{z}_o)}$$

$$- \frac{1}{2\pi i} \oint_{C_2} d\underline{z}' \frac{\underline{f}(\underline{z}')}{(\underline{z}' - \underline{z}_o) - (\underline{z} - \underline{z}_o)}, \tag{6.128}$$

and then generate quantities that can be expanded using the basic Taylor series of Equation 6.100:

$$\underline{f}(\underline{z}) = \frac{1}{2\pi i} \oint_{C_1} d\underline{z}' \frac{\underline{f}(\underline{z}')}{(\underline{z}' - \underline{z}_o)\left[1 - \frac{\underline{z}-\underline{z}_o}{\underline{z}'-\underline{z}_o}\right]}$$

$$+ \frac{1}{2\pi i} \oint_{C_2} d\underline{z}' \frac{\underline{f}(\underline{z}')}{(\underline{z} - \underline{z}_o)\left[1 - \frac{\underline{z}'-\underline{z}_o}{\underline{z}-\underline{z}_o}\right]}. \tag{6.129}$$

We have manipulated the two denominators differently in anticipation of the series expansions about to be performed. On C_1, $|\underline{z} - \underline{z}_o| < |\underline{z}' - \underline{z}_o|$, and we use the expansion

$$\frac{1}{1 - \frac{\underline{z}-\underline{z}_o}{\underline{z}'-\underline{z}_o}} = \sum_{n=0}^{\infty} \left(\frac{\underline{z} - \underline{z}_o}{\underline{z}' - \underline{z}_o}\right)^n. \tag{6.130}$$

On C_2, $|\underline{z} - \underline{z}_o| > |\underline{z}' - \underline{z}_o|$, so a different series is used:

$$\frac{1}{1 - \frac{\underline{z}'-\underline{z}_o}{\underline{z}-\underline{z}_o}} = \sum_{m=0}^{\infty} \left(\frac{\underline{z}' - \underline{z}_o}{\underline{z} - \underline{z}_o}\right)^m. \tag{6.131}$$

When these expansions are substituted into Equation 6.129, and the orders of the integrations and summations are reversed, the result is

$$\underline{f}(\underline{z}) = \sum_{n=0}^{\infty} \left[\frac{1}{2\pi i} \oint_{C_1} d\underline{z}' \frac{\underline{f}(\underline{z}')}{(\underline{z}' - \underline{z}_o)^{n+1}}\right] (\underline{z} - \underline{z}_o)^n$$

$$+ \sum_{m=0}^{\infty} \left[\frac{1}{2\pi i} \oint_{C_2} d\underline{z}' \underline{f}(\underline{z}')(\underline{z}' - \underline{z}_o)^m\right] \frac{1}{(\underline{z} - \underline{z}_o)^{m+1}}. \tag{6.132}$$

Equation 6.132 is not quite in the form we are seeking. The first summation, involving the integrals over C_1, is fine and can be identified with the $n \geq 0$ terms of the Laurent series. In the second summation, m goes from 0 to $+\infty$. If the substitution $m = -n-1$ is made, this summation can be identified with the $n < 0$ terms of a Laurent series.

Equation 6.132 can be rewritten as

$$f(z) = \sum_{n=0}^{\infty} \left[\frac{1}{2\pi i} \oint_{C_1} dz' \frac{f(z')}{(z' - z_o)^{n+1}} \right] (z - z_o)^n$$

$$+ \sum_{n=-1}^{-\infty} \left[\frac{1}{2\pi i} \oint_{C_2} dz' \frac{f(z')}{(z' - z_o)^{n+1}} \right] (z - z_o)^n. \qquad (6.133)$$

Comparing Equation 6.133 with Equation 6.125, the general expression for the Laurent series, the c_n for $n \geq 0$ are given by integrals over C_1

$$c_n = \frac{1}{2\pi i} \oint_{C_1} dz' \frac{f(z')}{(z' - z_o)^{n+1}} \quad \text{for } n \geq 0, \qquad (6.134)$$

while the c_n for $n < 0$ are given by integrals over C_2

$$c_n = \frac{1}{2\pi i} \oint_{C_2} dz' \frac{f(z')}{(z' - z_o)^{n+1}} \quad \text{for } n < 0. \qquad (6.135)$$

The Laurent series that is generated by these coefficients will be valid in the annular, crosshatched region of Figure 6.18. This region can be expanded by shrinking C_1 and expanding C_2, while keeping $f(z)$ analytic between the two circles. Therefore C_2 can be shrunk to the singularity circle passing through $z = a$, and C_1 can be expanded until it runs into another singularity of $f(z)$. If b is the first such singularity, the convergence region would be as shown in Figure 6.19. If $f(z)$ has no other singularities, C_2 can be expanded to infinity.

Once we have determined the maximum contour C_2 and minimum contour C_1, as described above, the coefficient expressions in Equations 6.134 and 6.135 can be combined into a single expression,

$$c_n = \frac{1}{2\pi i} \oint_C dz' \frac{f(z')}{(z' - z_o)^{n+1}} \quad \text{for all } n, \qquad (6.136)$$
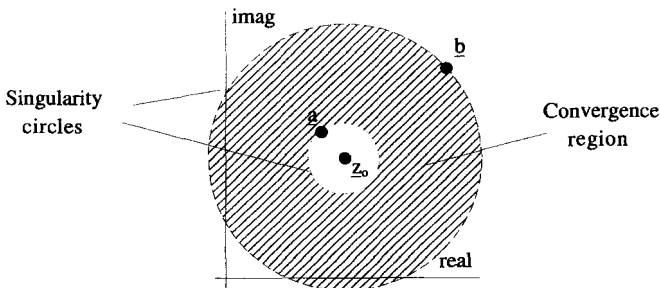

Figure 6.19 Convergence Region for the Laurent Series Expansion Between Two Nonanalytic Points
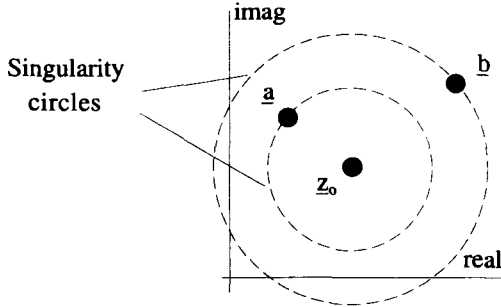
Figure 6.20   Complex Plane for Determining the Laurent Coefficients

where C is *any* contour which lies entirely within the two expanded contours. This last simplification is a direct consequence of the deformation theorem for analytic regions.

For $n < 0$, the only nonanalytic points of the integrand in Equation 6.136 are those associated with $\underline{f}(\underline{z})$. For $n \geq 0$ the integrand has singularities associated with $\underline{f}(\underline{z})$ as well as those generated by the $(\underline{z}' - \underline{z}_o)^{n+1}$ factor in the denominator. Notice that if $\underline{f}(\underline{z})$ is analytic everywhere inside C, all the $\underline{c}_n$ for $n < 0$ must be zero, and the Laurent series degenerates to a Taylor series, as expected.

Example 6.7   As an example of this process, we will determine the coefficients of the Laurent series expansion for the complex function

$$\underline{f}(\underline{z}) = \frac{1}{\underline{z} - \underline{a}} + \frac{1}{\underline{z} - \underline{b}} = \frac{2\underline{z} - \underline{a} - \underline{b}}{(\underline{z} - \underline{a})(\underline{z} - \underline{b})}. \tag{6.137}$$

This function is analytic everywhere in the complex plane, except at $\underline{z} = \underline{a}$ and $\underline{z} = \underline{b}$. The Laurent series is to be expanded around the point $\underline{z} = \underline{z}_o$. These points are shown in Figure 6.20, where we have assumed that $\underline{a}$ is closer to $\underline{z}_o$ than $\underline{b}$. For this example, the contour C used for determining the Laurent coefficients from Equation 6.136 must be centered at $\underline{z}_o$, but can be located in three general regions: inside the singularity circle of $\underline{a}$, between the $\underline{a}$ and $\underline{b}$ singularity circles, or outside the $\underline{b}$ singularity circle. These three situations will be treated separately.

The Contour C Inside the Small Singularity Circle   The case of the C contour lying inside the singularity circle of $\underline{a}$ is shown in Figure 6.21(a).The coefficients of the Laurent series are obtained from the Cauchy integration of Equation 6.136 as

$$\underline{c}_n = \frac{1}{2\pi i} \oint_C d\underline{z}' \frac{2\underline{z}' - \underline{a} - \underline{b}}{(\underline{z}' - \underline{a})(\underline{z}' - \underline{b})(\underline{z}' - \underline{z}_o)^{n+1}}. \tag{6.138}$$

For $n < 0$, the integrand of Equation 6.138 is analytic inside C and so for $n < 0$ all the $\underline{c}_n$ are zero. For $n \geq 0$, the only nonanalytic point of the integrand is at $\underline{z}_o$, and so
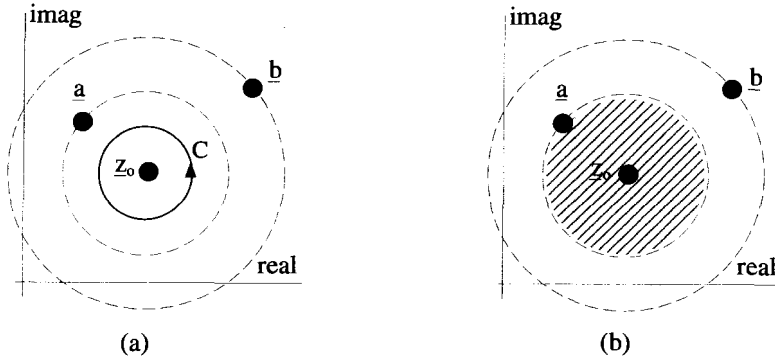
Figure 6.21   Complex Plane for C Inside the Small Singularity Circle

$$
\underline{c}_n = \frac{1}{n!} \frac{d^n}{d\underline{z}^n} \left\{ \frac{2\underline{z} - \underline{a} - \underline{b}}{(\underline{z} - \underline{a})(\underline{z} - \underline{b})} \right\} \Bigg|_{\underline{z} = \underline{z}_o}
$$

$$
= -\left(\frac{1}{\underline{a} - \underline{z}_o} \right)^{n+1} - \left(\frac{1}{\underline{b} - \underline{z}_o} \right)^{n+1} \quad n \geq 0. \tag{6.139}
$$

Therefore, for this case, the series expansion becomes

$$
\frac{2\underline{z} - \underline{a} - \underline{b}}{(\underline{z} - \underline{a})(\underline{z} - \underline{b})} = \sum_{n=0}^{\infty} -\left\{ \left(\frac{1}{\underline{a} - \underline{z}_o} \right)^{n+1} + \left(\frac{1}{\underline{b} - \underline{z}_o} \right)^{n+1} \right\} (\underline{z} - \underline{z}_o)^n. \tag{6.140}
$$

The region of convergence for this series can be determined by finding the largest and smallest circles around the point $\underline{z}_o$ that have $\underline{f}(\underline{z})$ analytic everywhere in between. In this case, the largest circle we can have lies just inside the singularity circle that passes through $\underline{a}$. Because there are no other nonanalytic points inside this circle, the smallest circle can be collapsed to a circle of radius zero. Thus the series converges for the entire crosshatched region shown in Figure 6.21(b).

In this case, it can be seen that the series in Equation 6.140 is not a Laurent series at all, but just a simple Taylor series with the standard Taylor series convergence region. This will always be the case if $\underline{f}(\underline{z})$ is analytic inside C.

The Contour C Between the Singularity Circles   Now let's look at the more interesting situation where the contour C lies between the singularity circles, as shown in Figure 6.22(a). The coefficients for the Laurent series can be obtained by Cauchy integration using an expression identical to Equation 6.138, except C is now between the singularity circles

$$
\underline{c}_n = \frac{1}{2\pi i} \oint_C d\underline{z}' \frac{2\underline{z}' - \underline{a} - \underline{b}}{(\underline{z}' - \underline{a})(\underline{z}' - \underline{b})(\underline{z}' - \underline{z}_o)^{n+1}}. \tag{6.141}
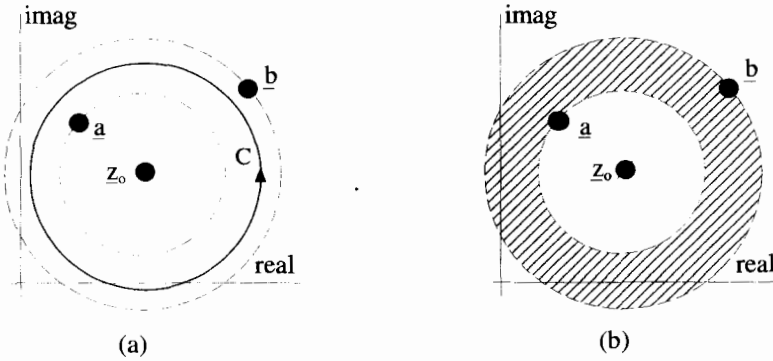$$

Figure 6.22   Complex Plane for C Between the Singularity Circles

For $n < 0$, the integrand of Equation 6.141 is no longer analytic inside C because of the singularity at $\underline{z} = \underline{a}$. Therefore,

$$\underline{c}_n = \left(\frac{1}{\underline{a} - \underline{z}_o}\right)^{n+1} \qquad \text{for } n < 0. \tag{6.142}$$

For $n \geq 0$, there are two singularities inside C, one at $\underline{z}_o$ and another at $\underline{a}$. Therefore, after some work, we obtain

$$\underline{c}_n = -\left(\frac{1}{\underline{b} - \underline{z}_o}\right)^{n+1} \qquad \text{for } n \geq 0. \tag{6.143}$$

For this case, the series expansion becomes

$$\frac{2\underline{z} - \underline{a} - \underline{b}}{(\underline{z} - \underline{a})(\underline{z} - \underline{b})} = \sum_{n=-\infty}^{-1} \left(\frac{1}{\underline{a} - \underline{z}_o}\right)^{n+1} (\underline{z} - \underline{z}_o)^n$$

$$+ \sum_{n=0}^{\infty} -\left(\frac{1}{\underline{b} - \underline{z}_o}\right)^{n+1} (\underline{z} - \underline{z}_o)^n. \tag{6.144}$$

The region of convergence for this series is obtained by determining the largest and smallest circles, centered around $\underline{z}_o$, that still have $\underline{f}(\underline{z})$ analytic everywhere in between. In this case, the largest circle is just inside the singularity circle passing through $\underline{b}$, while the smallest is the circle that lies just outside the singularity circle passing through $\underline{a}$. The resulting annular convergence region is shown in Figure 6.22(b).

In this case, we have a true Laurent series with $n < 0$ terms in the expansion. Notice how the terms for $n \geq 0$ make up the Taylor series expansion for the $1/(\underline{z} - \underline{b})$ part of $\underline{f}(\underline{z})$, while the $n < 0$ terms are a Laruent expansion for the $1/(\underline{z} - \underline{a})$ part. The $n < 0$ terms diverge inside the singularity circle that passes through $\underline{a}$, while

the $n \geq 0$ terms diverge outside the singularity circle that passes through $\underline{b}$. The combination only converges between the two singularity circles.

The Contour C Outside the Large Singularity Circle   As a final variation on this expansion, consider the situation where the contour C lies outside the large singularity circle that passes through $\underline{b}$, as shown in Figure 6.23(a). The coefficients for the expansion can be obtained by a Cauchy integration using an expression identical to Equation 6.138, except C is now outside the singularity circle that passes through $\underline{b}$:

$$\underline{c}_n = \frac{1}{2\pi i} \oint_C d\underline{z}' \, \frac{2\underline{z}' - \underline{a} - \underline{b}}{(\underline{z}' - \underline{a})(\underline{z}' - \underline{b})(\underline{z}' - \underline{z}_o)^{n+1}}. \tag{6.145}$$

For $n < 0$, the integrand of Equation 6.145 has two singularities, one at $\underline{z} = \underline{a}$ and the other at $\underline{z} = \underline{b}$, and we have

$$\underline{c}_n = \left(\frac{1}{\underline{a} - \underline{z}_o} \right)^{n+1} + \left(\frac{1}{\underline{b} - \underline{z}_o} \right)^{n+1} \qquad \text{for } n < 0. \tag{6.146}$$

For $n \geq 0$, there are three singularities inside C, at $\underline{z} = \underline{a}$, $\underline{z} = \underline{b}$, and $\underline{z} = \underline{z}_o$. After more work than last time, we find the negative coefficients are all zero:

$$\underline{c}_n = 0 \qquad \text{for } n \geq 0. \tag{6.147}$$

The series expansion, therefore, becomes

$$\frac{2\underline{z} - \underline{a} - \underline{b}}{(\underline{z} - \underline{a})(\underline{z} - \underline{b})} = \sum_{n=-\infty}^{-1} \left\{ \left(\frac{1}{\underline{a} - \underline{z}_o} \right)^{n+1} + \left(\frac{1}{\underline{b} - \underline{z}_o} \right)^{n+1} \right\} (\underline{z} - \underline{z}_o)^n. \tag{6.148}$$

As before, the region of convergence for this series is obtained by finding the largest and smallest circles centered at $\underline{z}_o$ that keep $\underline{f}(\underline{z})$ analytic between them. In this case
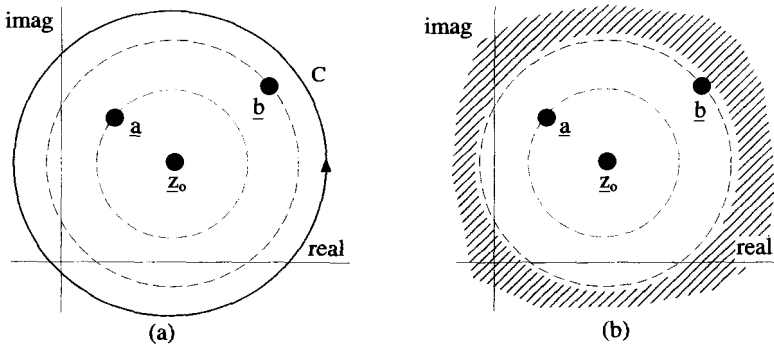

Figure 6.23   The Complex Plane for C Lying Outside the Large Singularity Circle

the smallest circle lies just outside the singularity circle that passes through $z = b$. The largest circle can be extended to infinity because there are no other nonanalytic points of $f(z)$. Therefore, the convergence region is as shown in Figure 6.23(b).

In this case we again obtain a true Laurent series. The coefficients of the series contain two terms. The first generates a Laurent series for $1/(z - a)$ and the second a Laurent series for $1/(z - b)$. The $1/(z - a)$ part of the series converges outside the singularity circle that passes through a. The $1/(z - b)$ part of the series converges outside the singularity circle that passes through b. The combination converges outside the largest of the two singularity circles, the one passing through b.

6.9.2   Laurent Coefficients Using the Basic Taylor Series

Evaluating series coefficients from Equation 6.136 is not always an easy task. Fortunately, like the Taylor series, Laurent series often can be generated from the basic series expansion of Equation 6.100. This is best demonstrated by an example.

Example 6.8   Consider the function

$$f(z) = \frac{1}{z - a}, \tag{6.149}$$

which is analytic at every point except $z = a$. We seek a Laurent series for this function expanded about $z_o = 0$ in the form

$$\frac{1}{z - a} = \sum_{n=-\infty}^{\infty} c_n z^n. \tag{6.150}$$

We can manipulate Equation 6.149 using Equation 6.100 to obtain

$$\frac{1}{z - a} = \left(\frac{1}{z}\right)\left(\frac{1}{1 - a/z}\right) = \left(\frac{1}{z}\right)\left[1 + \frac{a}{z} + \left(\frac{a}{z}\right)^2 + \cdots\right]. \tag{6.151}$$

This is in the form of a Laurent series with only $n < 0$ terms. The coefficients of Equation 6.150 are

$$c_n = \begin{cases} 0 & n \geq 0 \\ a^{-n-1} & n < 0 \end{cases}. \tag{6.152}$$

Because the series expansion in Equation 6.150 converges if

$$|z| > |a|, \tag{6.153}$$

this Laurent series is valid in the region shown in Figure 6.24. This is the region outside the singularity circle that passes through the point a. There is no outer limit to this convergence region.
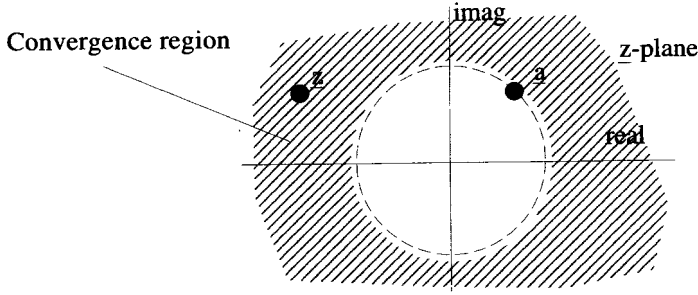
Figure 6.24 Convergence Region for Laurent Series Expansion About $z_o = 0$

The procedure for an expansion about an arbitrary point $z = z_o$ is similar. Manipulate Equation 6.149 as

$$\frac{1}{z - a} = \frac{1}{(z - z_o) - (a - z_o)}$$
$$= \left(\frac{1}{z - z_o}\right)\left(\frac{1}{1 - (a - z_o)/(z - z_o)}\right). \tag{6.154}$$

Expanding the second term gives

$$\frac{1}{z - a} = \left(\frac{1}{z - z_o}\right)\left[1 + \frac{a - z_o}{z - z_o} + \left(\frac{a - z_o}{z - z_o}\right)^2 + \cdots\right], \tag{6.155}$$

which again is a Laurent series with coefficients

$$c_n = \begin{cases} 0 & n \geq 0 \\ (a - z_o)^{-n-1} & n < 0 \end{cases}. \tag{6.156}$$

The series converges if

$$|z - z_o| > |a - z_o|, \tag{6.157}$$
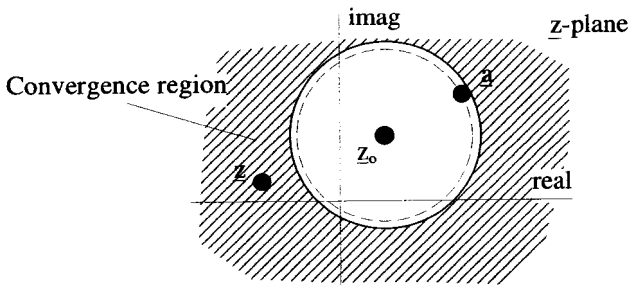
Figure 6.25 Convergence Region for Laurent Series Expanded about z_o

which is a region outside the singularity circle centered at z_o and passing through a, as shown in Figure 6.25. Notice, the region of convergence can be modified by changing the expansion point z_o. As z_o gets closer to a, the series becomes valid for more and more of the complex plane, but can never be made to converge exactly at $z = a$.

6.9.3 Classification of Singularities

Laurent expansions allow singularities of functions to be categorized. Imagine the point $z = a$ is a singularity of the function $f(z)$. If the most negative term in the Laurent series expansion around the singular point is $-m$, the singularity is called a *pole* of order m. For example, the function

$$f(z) = \frac{1}{z - a} \tag{6.158}$$

is its own Laurent expansion around $z = a$, and thus this function has a pole of order one at that point. This is often called a simple pole. The function

$$f(z) = \frac{1}{(z - a)^3} \tag{6.159}$$

again is its own Laurent Expansion around $z = a$ and therefore has a pole of order three.

It is possible to have a function that appears to have a pole, but on closer examination does not. An example is the function

$$f(z) = \frac{\sin(z)}{z}. \tag{6.160}$$

This function has a singularity at the point $z = 0$, because the denominator goes to zero, but the Laurent expansion around $z = 0$ has no negative terms. This type of singularity is called *removable* because the function does not diverge as we approach the singularity. This can be seen in the case of Equation 6.160 because

$$\lim_{z \to 0} \frac{\sin(z)}{z} = 1, \tag{6.161}$$

even though the value exactly at $z = 0$ is undefined.

It was quite easy to determine the order of the singularities in Equations 6.158 and 6.159, because they were in the convenient form of $1/(z - a)^m$. For a singularity of a more complicated function, you might think we would always have to determine the coefficients of the Laurent series. Fortunately, there is another approach, which is often much quicker. Notice for a pole of order m at $z = a$, multiplication by the factor $(z - a)^m$ will get rid of the divergence at a. In other words, if the original function

diverges because of a pole of order m at $z = a$, i.e.,

$$\lim_{z \to a} f(z) \to \pm \infty, \tag{6.162}$$

the divergence can be removed by multiplying by the factor $(z - a)^m$:

$$\lim_{z \to a} \left[f(z)(z - a)^m \right] \to \text{ a finite value.} \tag{6.163}$$

Thus, one way to determine the order of a pole is to multiply by successive factors of $(z - a)$, until the result no longer diverges at a. The number of multiplying factors needed determines the order of the pole.

A final type of singularity exists if the function diverges at $z = a$ and this divergence cannot be removed by multiplying by $(z - a)^m$ for any value of m. This is called an *essential* singularity. The Laurent expansion for a function about an essential singularity point will have an infinite number of negative n terms. The function $e^{1/z}$ has an essential singularity at $z = 0$. This function goes to infinity at $z = 0$ faster than any power of z goes to 0.

Example 6.9   To illustrate this technique, consider the function

$$f(z) = \frac{z}{\cos z - 1}. \tag{6.164}$$

There are singularities everywhere $\cos z = 1$. Let's consider just the singularity at $z = 0$.

First we will demonstrate that the function diverges as we approach $z = 0$. Notice that both the numerator and denominator are zero at this point, so we must use l'Hopital's rule to determine the limiting behavior of the function. The complex version of l'Hopital's rule is identical to the version for real functions:

$$\lim_{z \to z_o} \left[\frac{f_1(z)}{f_2(z)} \right] = \lim_{z \to z_o} \left[\frac{d f_1(z)/dz}{d f_2(z)/dz} \right] \tag{6.165}$$

as long as both

$$\lim_{z \to z_o} \left[f_1(z) \right] = 0$$

$$\lim_{z \to z_o} \left[f_2(z) \right] = 0. \tag{6.166}$$

Applying this to the $z = 0$ singularity of Equation 6.164 gives

$$\lim_{z \to 0} f(z) = \lim_{z \to 0} \left(\frac{z}{\cos z - 1} \right) = \lim_{z \to 0} \left(\frac{-1}{\sin z} \right) \to -\infty. \tag{6.167}$$

Now let's determine the order of the pole. Start by multiplying the function by $\underline{z}$ and taking the limit:

$$\lim_{\underline{z}\to 0}\left[f(\underline{z})\underline{z}\right] = \lim_{\underline{z}\to 0}\left[\left(\frac{\underline{z}}{\cos\underline{z}-1}\right)(\underline{z})\right] \to \frac{0}{0}. \tag{6.168}$$

To determine this limit, use l'Hopital's rule:

$$\lim_{\underline{z}\to 0}\left(\frac{\underline{z}^2}{\cos\underline{z}-1}\right) = \lim_{\underline{z}\to 0}\left[\frac{d(\underline{z}^2)/d\underline{z}}{d(\cos\underline{z}-1)/d\underline{z}}\right] = \lim_{\underline{z}\to 0}\left(\frac{-2\underline{z}}{\sin\underline{z}}\right) = \frac{0}{0} \tag{6.169}$$

not once, but twice:

$$\lim_{\underline{z}\to 0}\left(\frac{\underline{z}^2}{\cos\underline{z}-1}\right) = \lim_{\underline{z}\to 0}\left[\frac{d(-2\underline{z})/d\underline{z}}{d(\sin\underline{z})/d\underline{z}}\right] = \lim_{\underline{z}\to 0}\left(\frac{-2}{\cos\underline{z}}\right) = -2. \tag{6.170}$$

Multiplying by $\underline{z}$ has removed the divergence, so the singularity at $\underline{z}=0$ is a pole of order one.

6.10   THE RESIDUE THEOREM

It could be argued that the most useful result of this chapter is the Residue Theorem. This theorem provides a powerful tool for calculating both complex contour integrals and real, definite integrals.

6.10.1   The Residue of a Single Pole

Consider an integral of $\underline{w}(\underline{z})$ on an arbitrary closed contour in the $\underline{z}$-plane:

$$\underline{I} = \oint_C d\underline{z}\, w(\underline{z}). \tag{6.171}$$

Assume that $\underline{w}(\underline{z})$ has one singularity inside C, and that $\underline{w}(\underline{z})$ can be expanded in a Laurent series about a point $\underline{z}_o$, such that the contour C is entirely within the convergence region, as shown in Figure 6.26. Notice that a Laurent series is necessary because $\underline{w}(\underline{z})$ is not analytic everywhere inside C. Consequently, the series can be substituted for $\underline{w}(\underline{z})$ in the integral

$$\underline{I} = \oint_C d\underline{z}\, \sum_{n=-\infty}^{\infty} \underline{c}_n(\underline{z}-\underline{z}_o)^n. \tag{6.172}$$

Reversing the order of summation and integral gives

$$\underline{I} = \sum_{n=-\infty}^{\infty} \oint_C d\underline{z}\, \underline{c}_n(\underline{z}-\underline{z}_o)^n. \tag{6.173}$$
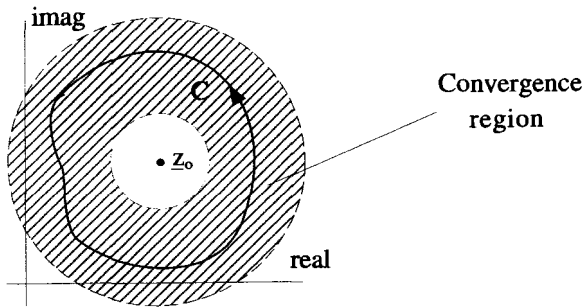
Figure 6.26   Contour and Convergence Annulus for Residue Theorem Integration

Consider any one of the terms with $n \geq 0$. Its contribution to I must be zero because the integrand is analytic everywhere:

$$\oint_C dz\, c_n (z - z_o)^n = 0 \qquad n \geq 0. \tag{6.174}$$

Now look at the $n = -2$ term. According to Cauchy's Integral Formula,

$$\oint_C dz \frac{c_{-2}}{(z - z_o)^2} = 2\pi i \left. \frac{dc_{-2}}{dz} \right|_{z=z_o} = 0, \tag{6.175}$$

so the contribution of this term to I is also zero. This will be the case for all terms of the series with $n \leq -2$. For $n = -1$, however,

$$I = \oint_C dz\, w(z) = \oint_C \frac{c_{-1}}{z - z_o} dz = 2\pi i c_{-1}, \tag{6.176}$$

where again, the last step follows from the Cauchy Integral Formula.

The c_{-1} coefficient is referred to as the *residue* at the singularity and, in general, the integral's value is

$$\oint_C dz\, w(z) = 2\pi i \times \text{ the residue.} \tag{6.177}$$

6.10.2   Residue Theorem for Integrals Surrounding Multiple Poles

The extension of the above arguments to cover functions with several poles is straightforward. Consider the complex function

$$w(z) = \frac{f(z)}{(z - a)(z - b)(z - c)}, \tag{6.178}$$
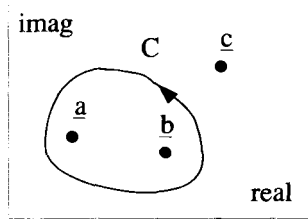
Figure 6.27   The Residue Theorem Applied to Multipoled Functions

where $f(z)$ is analytic in the entire z-plane. There are three poles at $z = a, z = b$, and $z = c$. Imagine we are interested in evaluating an integral

$$I = \oint_C dz\, w(z), \tag{6.179}$$

where C encloses two of the poles of $w(z)$, as shown in Figure 6.27. Because the Cauchy Integral Theorem allows us to arbitrarily deform the contour through analytic regions, the integral can be broken up into the sum of two individual integrals, one around each pole:

$$I = \oint_C dz\, w(z) = \oint_{C_1} dz\, w(z) + \oint_{C_2} dz\, w(z), \tag{6.180}$$

as shown in Figure 6.28. Therefore, the value of the integral will be $2\pi i$ times the sum of the residues of $w(z)$ at a and b:

$$\oint_C \frac{dz\, f(z)}{(z - a)(z - b)(z - c)} = 2\pi i[\text{residue at } a + \text{residue at } b]. \tag{6.181}$$

The general form of the Residue Theorem for a function with an arbitrary number of poles is

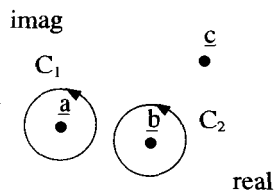$$\oint_C dz\, w(z) = 2\pi i \sum \text{Residues of poles inside } C. \tag{6.182}$$

Figure 6.28   Equivalent Multipole Contours

6.10.3   Residue Formulae

Now that we have shown why residues are important, we need to develop some methods for calculating them. We showed earlier that the residue of a singularity is just the c_{-1} term in a Laurent series expanded around the singularity itself. Thus, we can always determine the residue of a singularity once we know the Laurent series. Sometimes this is the only way possible, but in many cases we don't have to do that much work.

Imagine we already know that a singularity at $z = a$ is a first-order pole. The Laurent series for such a pole, expanded about the pole, looks like this,

$$w(z) = \frac{c_{-1}}{(z - a)} + c_o + c_1(z - a) + c_2(z - a)^2 + \cdots, \qquad (6.183)$$

with no terms for $n < -1$. Multiplying this series by $(z - a)$ gives

$$w(z)(z - a) = c_{-1} + c_o(z - a) + c_1(z - a)^2 + c_2(z - a)^3 + \cdots. \qquad (6.184)$$

If then we take the limit as $z \to a$, we find that

$$\lim_{z \to a}[w(z)(z - a)] = c_{-1}. \qquad (6.185)$$

This gives a very simple method for determining the residue of a first-order pole. As an example, consider the function

$$w(z) = \frac{\sin z}{(z - a)(z - b)}, \qquad (6.186)$$

which has first-order poles at both $z = a$ and $z = b$. Let's calculate the residue at $z = a$:

$$\lim_{z \to a}[w(z)(z - a)] = \lim_{z \to a}\left[\frac{\sin z}{(z - a)(z - b)}(z - a)\right] = \frac{\sin a}{(a - b)}. \qquad (6.187)$$

This method can be extended for the residues of higher-order poles. The Laurent expansion of a second-order pole located at $z = a$, expanded around the pole, has the form

$$w(z) = \frac{c_{-2}}{(z - a)^2} + \frac{c_{-1}}{(z - a)} + c_o + c_1(z - a) + \cdots. \qquad (6.188)$$

Multiplication by $(z - a)^2$ gives

$$w(z)(z - a)^2 = c_{-2} + c_{-1}(z - a) + c_o(z - a)^2 + \cdots. \qquad (6.189)$$

Now we take the derivative

$$\frac{d}{dz}\left[w(z)(z - a)^2\right] = c_{-1} + 2c_0(z - a) + 3c_1(z - a)^2 + \cdots, \qquad (6.190)$$

and finally take the limit as $\underline{z} \to \underline{a}$:

$$\lim_{\underline{z} \to \underline{a}} \frac{d}{d\underline{z}} \left[w(\underline{z})(\underline{z} - \underline{a})^2 \right] = \underline{c}_{-1}. \tag{6.191}$$

This gives us a simple formula for calculating the residue of a second-order pole. For a pole of order n located at $\underline{z} = \underline{a}$, Equation 6.191 generalizes to

$$\underline{c}_{-1} = \frac{1}{(n-1)!} \lim_{\underline{z} \to \underline{a}} \frac{d^{n-1}}{d\underline{z}^{n-1}} \left[w(\underline{z})(\underline{z} - \underline{a})^n \right]. \tag{6.192}$$

6.11 DEFINITE INTEGRALS AND CLOSURE

The residue theorem can be used to evaluate certain types of real, definite integrals which appear frequently in physics and engineering problems. The method involves converting a real integral into a complex contour integral and then *closing* the contour. The residue theorem can then be used to evaluate the contour integral, and subsequently the original definite integral.

Let x be a pure real variable, and consider an integral with the form

$$I = \int_{-\infty}^{+\infty} dx \, w(x), \tag{6.193}$$

where $w(x)$ is a real function of x. Even though this integral involves quantities that are all pure real, it can be evaluated with x represented by a complex variable $\underline{z}$ by performing the integration along the real axis of the complex plane. That is

$$I = \int_{-\infty}^{+\infty} dx \, w(x) = \int_R d\underline{z} \, w(\underline{z}), \tag{6.194}$$

where R is a straight-line path along the real axis from $\underline{z} = -\infty$ to $\underline{z} = +\infty$, as shown in Figure 6.29. To proceed, we need to define the function $\underline{w}(\underline{z})$ for values of $\underline{z}$ off the real axis. We can do this with a process called *analytic continuation*. Basically, this involves the determination of a complex function which, along the real axis, is the same as the original real function, and is analytic in as much of the complex plane as possible. In general, the continuation of a real function can be performed by
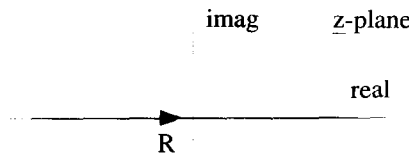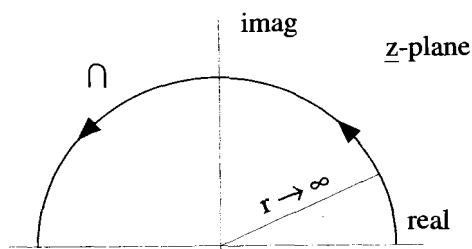

Figure 6.29 Equivalent Path for Real Integral

Figure 6.30   Semicircular Contour for Closure

simply replacing all the x's in a function with complex $\underline{z}$'s. For example, the analytic continuation of $w(x) = x^2$ would be $\underline{w}(\underline{z}) = \underline{z}^2$.

Because R is not a closed path, the residue theorem cannot be used to perform the complex integration of Equation 6.194. However, suppose

$$\int_{\cap} d\underline{z}\,\underline{w}(\underline{z}) = 0, \tag{6.195}$$

where $\cap$ is a semicircular path of infinite radius in the upper half $\underline{z}$-plane, as shown in Figure 6.30. Because this additional section cannot change the value of the contour integral, we can just add it to the original complex integral to form a closed contour. That is, if the integral along $\cap$ is zero, then

$$I = \int_R d\underline{z}\,w(\underline{z}) = \int_R d\underline{z}\,w(\underline{z}) + \int_{\cap} d\underline{z}\,\underline{w}(\underline{z}). \tag{6.196}$$

The combination of the R and $\cap$ paths form the closed contour C shown in Figure 6.31. The original definite, real integral can now be written as a closed integral that can be evaluated by using the residue theorem:

$$\int_{-\infty}^{+\infty} dx\,w(x) = \oint_C d\underline{z}\,\underline{w}(\underline{z})$$

$$= 2\pi i \sum \text{ Residues of } \underline{w}(\underline{z}) \text{ inside } C. \tag{6.197}$$
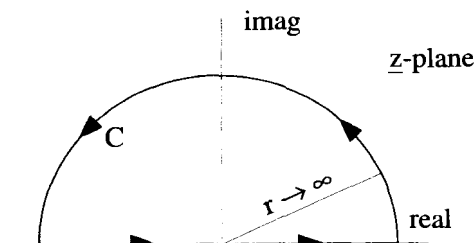

Figure 6.31   Closed Contour for Definite Integral

The technique of adding a section of contour to form a closed contour is referred to as closure. In the discussion above, closure was accomplished by adding a counterclockwise, infinite radius semicircular path in the upper half plane. Sometimes closure is accomplished by a similar semicircular path in the lower half plane. Closure is most often performed with semicircles, but there are cases where rectangles or other shapes are necessary.

6.11.1   Conditions for Closure on a Semicircular Path

As mentioned, for Equation 6.197 to be valid, the integral along $\cap$ must be zero. This integration can be viewed as a semicircular contour, as shown in Figure 6.30, in the limit of the radius approaching infinity. If we let r be the radius of the semicircle, then on $\cap$

$$\underline{z} = re^{i\theta}, \tag{6.198}$$

where $0 < \theta < \pi$ for a semicircle in the upper complex plane. To move along $\cap$,

$$d\underline{z} = ire^{i\theta}d\theta, \tag{6.199}$$

and the integral on $\cap$ becomes

$$\int_{\cap} d\underline{z}\,\underline{w}(\underline{z}) = \lim_{r \to \infty} \int_0^\pi d\theta\, ire^{i\theta}\underline{w}(re^{i\theta}). \tag{6.200}$$

Thus to successfully use closure in the form described in the previous section, $\underline{w}(\underline{z})$ must obey

$$\lim_{r \to \infty} \int_0^\pi d\theta\, ire^{i\theta}\underline{w}(re^{i\theta}) = 0. \tag{6.201}$$

Using Equation 6.201 can be a little unwieldy, but a more usable condition can be derived by using the inequality

$$\left| \int d\underline{z}\,\underline{f}(\underline{z}) \right| \le \int |d\underline{z}\,\underline{f}(\underline{z})|, \tag{6.202}$$

which holds for any complex integral. Applying this to Equation 6.201, and noting $|e^{i\theta}| = 1$, gives

$$\lim_{r \to \infty} \left| \int_0^\pi d\theta\, ire^{i\theta}\underline{w}(re^{i\theta}) \right| \le \lim_{r \to \infty} \int_0^\pi |d\theta\, r\underline{w}(re^{i\theta})|$$

$$\le \pi \lim_{r \to \infty} \{rW_{max}\}, \tag{6.203}$$

where W_{max} is the largest value of $|\underline{w}(\underline{z})|$ on the interval $0 < \theta < \pi$. This gives us a useful result. If W_{max} shrinks to zero faster than $1/r$, then the contribution of the $\cap$

contour necessarily must vanish. Notice, it is not enough that $|\underline{w}(\underline{z})| \to 0$ as $r \to \infty$. Because of the extra factor of r, $|\underline{w}(\underline{z})|$ must go to zero faster than $1/r$.

Example 6.10   As an example of this closure technique, consider the real integral,

$$I = \int_{-\infty}^{\infty} \frac{dx}{x^2 + 1}. \tag{6.204}$$

The value of the integral is equal to the area under the curve shown in Figure 6.32 and is clearly a positive, real quantity. It is easy to extend the integral into the complex plane

$$I = \int_R \frac{d\underline{z}}{\underline{z}^2 + 1}, \tag{6.205}$$

where the contour R is along the real $\underline{z}$-axis, as shown in Figure 6.29. Now consider the $\cap$ integral of Figure 6.30:

$$I_\cap = \int_\cap \frac{d\underline{z}}{\underline{z}^2 + 1}. \tag{6.206}$$

On $\cap$,

$$\underline{z} = re^{i\theta} \tag{6.207}$$

$$d\underline{z} = ire^{i\theta} d\theta \tag{6.208}$$

and

$$\frac{1}{\underline{z}^2 + 1} = \frac{1}{r^2 e^{i2\theta} + 1}$$

$$= \frac{1}{r^2} \left(\frac{1}{e^{i2\theta} + 1/r^2} \right). \tag{6.209}$$
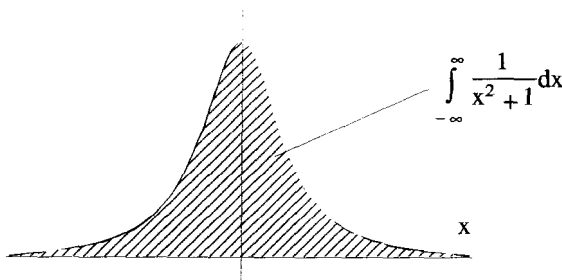
Figure 6.32   Real Integral to Be Evaluated by Closure
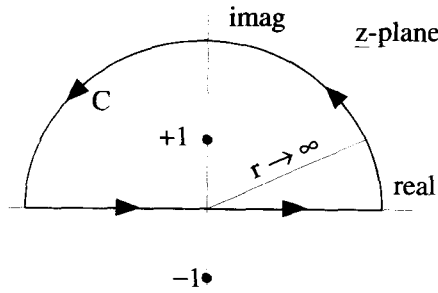
Figure 6.33   The Closed Contour for the Integration of $1/(x^2 + 1)$

This falls off faster than $1/r$ for all values of θ, so in the limit as $r \to \infty$, $I_\cap = 0$, and the original integral is equivalent to

$$\int_{-\infty}^{\infty} \frac{dx}{x^2 + 1} = \int_R \frac{dz}{z^2 + 1} + \int_\cap \frac{dz}{z^2 + 1}$$

$$= \oint_C \frac{dz}{z^2 + 1}, \tag{6.210}$$

where C is the closed contour shown in Figure 6.33. The two poles of the integrand at $\pm i$ have also been indicated in this figure. From the residue theorem, the value of the original integral must be

$$\int_{-\infty}^{\infty} \frac{dx}{x^2 + 1} = 2\pi i \sum \text{Residues of} \frac{1}{z^2 + 1} \text{ inside } C. \tag{6.211}$$

Only the $z = i$ pole is inside C, and a quick calculation shows the residue there is $1/2i$. Thus we obtain the final result

$$\int_{-\infty}^{\infty} \frac{dx}{x^2 + 1} = 2\pi i \left(\frac{1}{2i} \right) = \pi. \tag{6.212}$$

This particular integral could also have been closed with a semicircle in the lower half of the complex plane, as shown in Figure 6.34. In this figure, we see that now the $z = -i$ pole is enclosed and

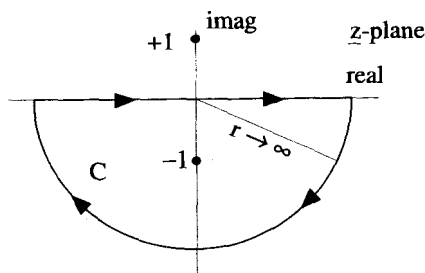$$\int_{-\infty}^{\infty} \frac{dx}{x^2 + 1} = -2\pi i (\text{Residue at } -i). \tag{6.213}$$

Figure 6.34   An Equivalent Closed Contour for the Integration of $1/(x^2 + 1)$

There is a minus sign out front because we enclosed the pole in a clockwise sense. The residue of this pole is $-1/2i$ so the value of the integral becomes

$$\int_{-\infty}^{\infty} \frac{dx}{x^2 + 1} = -2\pi i \left(\frac{-1}{2i} \right) = \pi, \tag{6.214}$$

which, fortunately, is the same as before.

Not all real integrals can be closed in both the upper and lower half-plane, as this one could. Some functions diverge in one half-plane and thus force the closing contour to be drawn in the other. Some nasty functions diverge on both sides, and make evaluation with this type of contour impossible.

6.11.2   Closure with Nonzero Contributions

In some cases we can get away with closing a contour, even if the integral along the added segment is not zero. As before, we extend the real integral into the complex plane. When we evaluate the integral, however, we must subtract off the added contribution,

$$\int_{-\infty}^{\infty} dx\, w(x) = \int_R d\underline{z}\, w(\underline{z}) = \oint_C d\underline{z}\, w(\underline{z}) - \int_{\cap} d\underline{z}\, w(\underline{z}), \tag{6.215}$$

where C is a closed contour, which includes both the path R along the real axis and the closing $\cap$ path.

Example 6.11   As an example, consider the definite integral

$$I = \int_{-\infty}^{\infty} dx\, \frac{e^{x/2}}{1 + \cosh x}. \tag{6.216}$$

As usual, the conversion to a complex integral is simple:

$$\int_{-\infty}^{\infty} dx\, \frac{e^{x/2}}{1 + \cosh x} = \int_R d\underline{z}\, \frac{e^{\underline{z}/2}}{1 + \cosh \underline{z}}. \tag{6.217}$$

If we try to close the contour with an infinite radius semicircle on either side of the real axis, we run into problems, because the complex integrand does not vanish as the radius approaches infinity. This can be seen most easily by evaluating the integrand along the imaginary axis, i.e., $\underline{z} = iy$, where we find

$$\left. \frac{e^{\underline{z}/2}}{1 + \cosh \underline{z}} \right|_{\underline{z}=iy} = \frac{2e^{iy/2}}{2 + e^{iy} + e^{-iy}}. \tag{6.218}$$

Notice as $y \to \pm\infty$, the RHS of Equation 6.218 oscillates. Therefore, closing the contour with a semicircular loop as we did in the previous section will not work.

By examining the poles of this function, you can see another big problem with using the infinite semicircular contour. A pole occurs every place where $\underline{z} = \pm(2n + 1)i\pi$ for all integer values of n. That means there are an infinite number of poles along the imaginary axis which would be enclosed by such a contour.

A better method is to close the contour as shown in Figure 6.35, and then take the limit $x_o \to \infty$. This contour encloses only a single pole, the one at $\underline{z} = i\pi$. The integral around the contour C, in this limit, has four parts:

$$\oint_C d\underline{z} \frac{e^{\underline{z}/2}}{1 + \cosh \underline{z}} d\underline{z} = \lim_{x_o \to \infty} \left\{ \int_{-x_o}^{x_o} d\underline{z} \frac{e^{\underline{z}/2}}{1 + \cosh \underline{z}} \right.$$

$$+ \int_{x_o}^{x_o + i2\pi} d\underline{z} \frac{e^{\underline{z}/2}}{1 + \cosh \underline{z}}$$

$$+ \int_{x_o + i2\pi}^{-x_o + i2\pi} d\underline{z} \frac{e^{\underline{z}/2}}{1 + \cosh \underline{z}} \tag{6.219}$$

$$\left. + \int_{-x_o + i2\pi}^{-x_o} d\underline{z} \frac{e^{\underline{z}/2}}{1 + \cosh \underline{z}} \right\}.$$

For the right vertical segment, $\underline{z} = x_o + iy$ and $d\underline{z} = idy$, so

$$\int_{x_o}^{x_o + i2\pi} d\underline{z} \frac{e^{\underline{z}/2}}{1 + \cosh \underline{z}} = \int_0^{2\pi} dy \frac{2ie^{x_o/2}e^{iy/2}}{2 + e^{x_o}e^{iy} + e^{-x_o}e^{-iy}}. \tag{6.220}$$
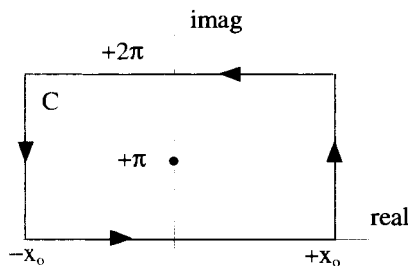

Figure 6.35 Contour for Integration of $e^{x/2}/(1 + \cosh x)$

As $x_o \to +\infty$, the e^{x_o} term in the denominator dominates and forces this integral to zero. The same thing happens for left vertical segment. The integral along the top, horizontal segment (from $x_o + i2\pi$ to $-x_o + i2\pi$) can be expressed in terms of the integral along the bottom segment:

$$\int_{x_o+i2\pi}^{-x_o+i2\pi} d\underline{z}\, \frac{e^{\underline{z}/2}}{1+\cosh\underline{z}} = \int_{x_o}^{-x_o} dx\, \frac{2e^{x/2}e^{i2\pi/2}}{2 + e^x e^{i2\pi} + e^{-x}e^{-i2\pi}}$$

$$= \int_{-x_o}^{x_o} dx\, \frac{2e^{x/2}}{2 + e^x + e^{-x}} \qquad (6.221)$$

$$= \int_{-x_o}^{x_o} d\underline{z}\, \frac{e^{\underline{z}/2}}{1+\cosh\underline{z}}.$$

Consequently,

$$\oint_C d\underline{z}\, \frac{e^{\underline{z}/2}}{1+\cosh\underline{z}} = 2 \lim_{x_o \to \infty} \int_{-x_o}^{x_o} d\underline{z}\, \frac{e^{\underline{z}/2}}{1+\cosh\underline{z}}. \qquad (6.222)$$

The right side of Equation 6.222 is just twice the real integral we started with, so

$$\int_{-\infty}^{\infty} dx\, \frac{e^{x/2}}{1+\cosh x} = \frac{1}{2} \oint_C d\underline{z}\, \frac{e^{\underline{z}/2}}{1+\cosh\underline{z}}. \qquad (6.223)$$

The contour C encloses the singularity at $i\pi$, as shown in Figure 6.35. The residue of the integrand at $i\pi$ can be determined by the techniques described earlier in this chapter. In this case, however, it may be simpler to look at the Taylor series for $\cosh\underline{z}$ expanded about $i\pi$. Using Equation 6.109, the first few terms of the expansion are

$$\cosh\underline{z} = -1 - \frac{1}{2}(\underline{z} - i\pi)^2 - \frac{1}{4!}(\underline{z} - i\pi)^4 + \cdots. \qquad (6.224)$$

This series converges to $\cosh\underline{z}$ in the entire $\underline{z}$-plane. From this Taylor series, you can infer that the complex integrand in Equation 6.223 has a second order pole at $i\pi$, because as we move arbitrarily close to this point, the denominator of the integrand is dominated by the $(\underline{z} - i\pi)^2$ term in the series expansion. Equation 6.191 allows us to calculate the residue as

$$\frac{d}{d\underline{z}}(-2e^{\underline{z}/2})\Big|_{i\pi} = -i. \qquad (6.225)$$

Therefore, according to Equation 6.223, the value of the real integral is

$$\int_{-\infty}^{\infty} dx\, \frac{e^{x/2}}{1+\cosh x} = \frac{1}{2}(2\pi) = \pi. \qquad (6.226)$$
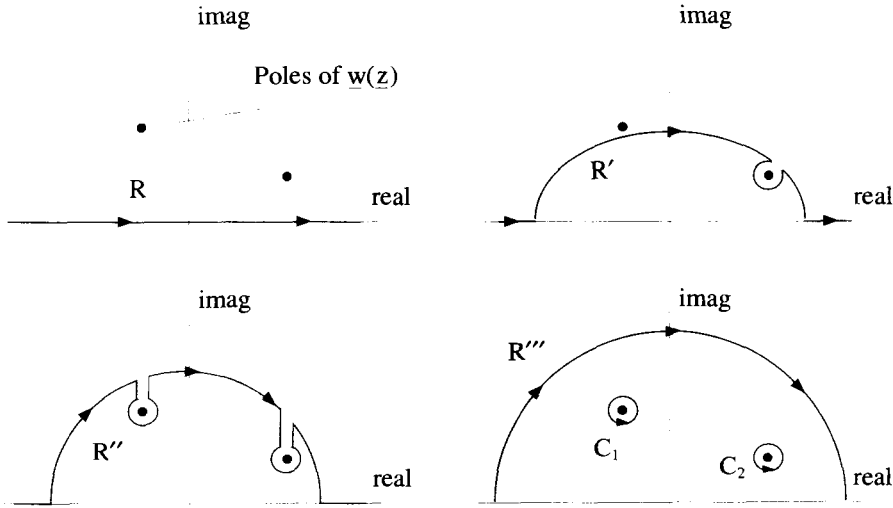
imag   imag

Poles of $\underline{w}(\underline{z})$

Figure 6.36   Closure Results from Contour Deformation

6.11.3   Another View of the Closure Process

There is an insightful, alternative viewpoint one can take when looking at closure problems. Instead of adding a contour segment to close an integral, the original contour, which lies along the real axis, can be deformed as shown in the sequence of Figure 6.36.

Consider a real integral

$$I = \int_{-\infty}^{\infty} dx\, w(x). \tag{6.227}$$

This integral converts to an equivalent complex integral along R, the real $\underline{z}$-axis,

$$\int_{-\infty}^{\infty} dx\, w(x) = \int_{R} d\underline{z}\, w(\underline{z}), \tag{6.228}$$

as shown in the upper left-hand corner of Figure 6.36. Two poles of $\underline{w}(\underline{z})$ are also shown in this figure. The contour R can be deformed without changing the value of the integral as long as the deformation does not cross a nonanalytic point of $\underline{w}(\underline{z})$. Therefore the integrals on R' and R'', shown in the upper right and lower left of Figure 6.36, are all equivalent to the original integral

$$\int_{-\infty}^{\infty} dx\, w(x) = \int_{R} \underline{w}(\underline{z})\, d\underline{z} = \int_{R'} d\underline{z}\, \underline{w}(\underline{z}) = \int_{R''} d\underline{z}\, \underline{w}(\underline{z}). \tag{6.229}$$

We can now make the vertical line segments of R'' overlap so they exactly cancel, as shown in the bottom right of Figure 6.36. The original integral can then be expressed as

$$I = \int_{-\infty}^{\infty} dx\, w(x) = \int_{R'''} d\underline{z}\, w(\underline{z}) + 2\pi i \sum \text{residues of poles crossed.} \tag{6.230}$$

If there are no other poles of $\underline{w}(\underline{z})$ then R''' can be expanded to the infinite semicircle we used before. If the contribution along R''' goes to zero as it is deformed to infinity, the result is the one found earlier in Equation 6.197. If, on the other hand, the contribution is nonzero, we must take that into account as we did in the previous section.

6.11.4 The Principal Part of an Integral

When the integrand of a real integral has a singularity for some value of x over the range of integration, the value of the integral is undefined. For instance the integrand of

$$\int_{-\infty}^{\infty} \frac{dx}{x^3} \tag{6.231}$$

behaves this way. However, since $1/x^3$ is an odd function of x, it is tempting to claim that the value of this integral is zero, because the negative infinity on the left of $x = 0$ is counteracted by the positive infinity on the right of $x = 0$. We can rescue our intuitive feel for this integral by defining the principal part of the integral:

$$PP \int_{-\infty}^{\infty} \frac{dx}{x^3} \equiv \lim_{\epsilon \to 0} \left\{ \int_{-\infty}^{-\epsilon} \frac{dx}{x^3} + \int_{\epsilon}^{\infty} \frac{dx}{x^3} \right\}. \tag{6.232}$$

In essence, the principal part takes the original integral and removes an infinitesimal region around the singularity. Evaluation of Equation 6.232 is straightforward:

$$\int_{-\infty}^{-\epsilon} \frac{dx}{x^3} + \int_{\epsilon}^{\infty} \frac{dx}{x^3} = \left. \frac{-1}{2x^2} \right|_{-\infty}^{-\epsilon} + \left. \frac{-1}{2x^2} \right|_{\epsilon}^{\infty}$$

$$= \frac{-1}{2\epsilon^2} + \frac{1}{2\epsilon^2}$$

$$= 0. \tag{6.233}$$

So, we in fact obtain the intuitive result

$$PP \int_{-\infty}^{\infty} \frac{dx}{x^3} = 0. \tag{6.234}$$

The principal part of many real integrals can be evaluated by closure and the residue theorem. Consider a real function $w(x)$ that blows up at $x = a$. The principal part of this integral is defined as

$$PP \int_{-\infty}^{\infty} dx\, w(x) \equiv \lim_{\epsilon \to 0} \left\{ \int_{-\infty}^{a-\epsilon} dx\, w(x) + \int_{a+\epsilon}^{\infty} dx\, w(x) \right\}. \tag{6.235}$$
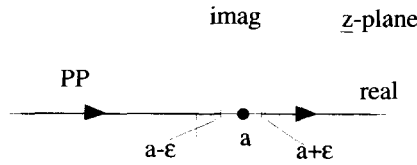
Figure 6.37 Principal Part Integration in the $\underline{z}$-Plane

This is equivalent to an integral done in the complex $\underline{z}$ plane,

$$PP \int_{-\infty}^{\infty} dx\, w(x) = \int_{PP} d\underline{z}\, w(\underline{z}), \tag{6.236}$$

where PP is a path of integration along the real $\underline{z}$-axis from $(-\infty, a - \epsilon)$ and $(a + \epsilon, \infty)$, in the limit of $\epsilon \to 0$, as shown in Figure 6.37. This integration can be handled with closure techniques by adding the standard infinite radius semicircular path in either the upper or lower complex plane, plus a semicircular path of radius ϵ around the singularity at $\underline{z} = a$. In other words, we write

$$\oint_C d\underline{z}\, w(\underline{z}) = \int_{PP} d\underline{z}\, w(\underline{z}) + \int_{\cup} d\underline{z}\, \underline{w}(\underline{z}) + \int_{\cap} d\underline{z}\, \underline{w}(\underline{z}), \tag{6.237}$$

where $\cup$ is the semicircular path of radius ϵ, $\cap$ is the semicircular path of infinite radius, and C is the closed contour made up of $PP + \cup + \cap$, as shown in Figure 6.38.

The evaluation of the closed integral over the contour C is simply equal to $2\pi i$ times the sum of the residues of $\underline{w}(\underline{z})$ inside C. Most often, the contribution from $\cap$ is zero. The integral over $\cup$ is a bit more tricky. If the singularity on the contour is a first-order pole, this added contribution is always half the residue of the pole. This is best demonstrated with an example.
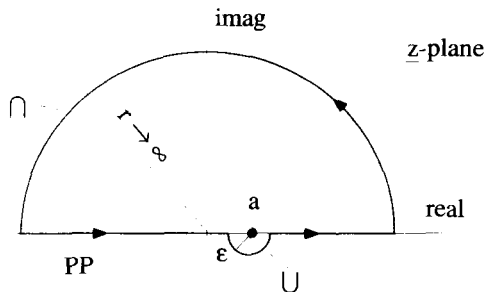
Figure 6.38 Closed Contour for Evaluating the Principal Part

Example 6.12   Let's evaluate

$$PP \int_{-\infty}^{\infty} dx \, \frac{1}{(x^2 + 1)(x - 1)} \equiv \lim_{\epsilon \to 0} \left\{ \int_{-\infty}^{1-\epsilon} dx \, \frac{1}{(x^2 + 1)(x - 1)} \right.$$

$$\left. + \int_{1+\epsilon}^{\infty} dx \, \frac{1}{(x^2 + 1)(x - 1)} \right\}. \qquad (6.238)$$

Applying the above arguments,

$$PP \int_{-\infty}^{\infty} dx \, \frac{1}{(x^2 + 1)(x - 1)} = \oint_C dz \, \frac{1}{(z^2 + 1)(z - 1)}$$

$$- \int_{\cup} dz \, \frac{1}{(z^2 + 1)(z - 1)}$$

$$- \int_{\cap} dz \, \frac{1}{(z^2 + 1)(z - 1)}, \qquad (6.239)$$

where the poles of the integrand and the various paths of integration are shown in Figure 6.39.

The contribution from $\cap$ goes to zero because, for large $|z|$, the integrand falls off as $1/z^3$. The closed integral on C is

$$\oint_C dz \, \frac{1}{(z^2 + 1)(z - 1)} = 2\pi i \sum \text{residues inside } C. \qquad (6.240)$$

There are two enclosed poles, one at $z = 1$, with a residue of $1/2$, and the other at $z = i$, with a residue of $(i - 1)/4$. Therefore,

$$\oint_C dz \, \frac{1}{(z^2 + 1)(z - 1)} = 2\pi i \left[\frac{1}{2} + \frac{i - 1}{4} \right] = \frac{\pi}{2}(i - 1). \qquad (6.241)$$
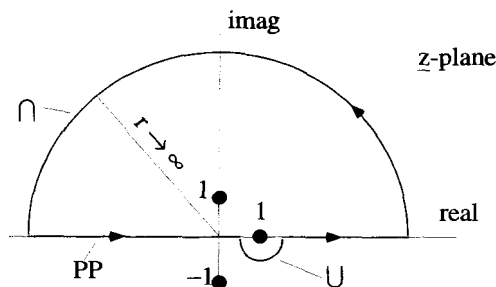
Figure 6.39   Contours for PP Integration of $1/(x^2 + 1)(x - 1)$
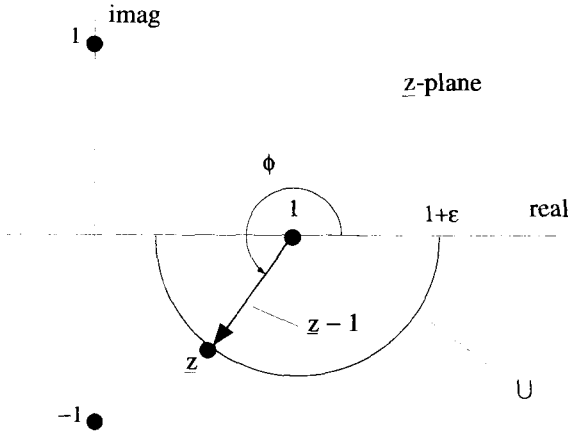
Figure 6.40   Expanded View of the ∪ Integration

The integration over the small semicircle ∪ needs particular attention. Consider the expanded view shown in Figure 6.40 where we have drawn a phasor $(\underline{z}-1) = \epsilon e^{i\phi}$. Along the path, $d\underline{z} = i e^{i\phi} d\phi$, and ϕ ranges from π to 2π. In the limit as $\epsilon \to 0$, we have

$$\lim_{\epsilon \to 0} \frac{1}{\underline{z}^2 + 1} = 2, \tag{6.242}$$

so the integral along ∪ becomes

$$\int_{\cup} d\underline{z} \frac{1}{(\underline{z}^2 + 1)(\underline{z} - 1)} = \lim_{\epsilon \to 0} \int_{\pi}^{2\pi} d\phi \frac{i\epsilon e^{i\phi}}{2\epsilon e^{i\phi}}$$

$$= \frac{i\pi}{2}. \tag{6.243}$$

The value of the principal part of the integral therefore becomes

$$PP \int_{-\infty}^{\infty} dx \frac{1}{(x^2 + 1)(x - 1)} = \oint_C d\underline{z} \frac{1}{(\underline{z}^2 + 1)(\underline{z} - 1)}$$

$$- \int_{\cup} d\underline{z} \frac{1}{(\underline{z}^2 + 1)(\underline{z} - 1)}$$

$$= -\frac{\pi}{2}. \tag{6.244}$$

Note that the principal part in the example could have been evaluated with contours other than those shown in Figure 6.39. The semicircle of radius ϵ could have been

taken above the singularity at $\underline{z} = 1$, and/or the infinite radius contour could have been placed in the lower half of the $\underline{z}$-plane. The result obtained for the principal part of the integral is the same with any of these choices.

6.11.5 Definite Integrals with $\sin\theta$ and $\cos\theta$

Real integrals of the form

$$I = \int_0^{2\pi} d\theta\, w(\sin\theta, \cos\theta), \tag{6.245}$$

where the integrand is a function of $\sin\theta$ and $\cos\theta$, can sometimes be converted to complex integrals along a circular contour of unit radius, as shown in Figure 6.41. On this contour, $\underline{z} = e^{i\theta}$ and θ ranges from 0 to 2π. Therefore, on C we can write

$$d\theta = \frac{d\underline{z}}{i\underline{z}} \tag{6.246}$$

and

$$\begin{aligned}
\sin\theta &= \frac{1}{2i}(e^{i\theta} - e^{-i\theta}) = \frac{1}{2i}(\underline{z} - 1/\underline{z}) \\
\cos\theta &= \frac{1}{2}(e^{i\theta} + e^{-i\theta}) = \frac{1}{2}(\underline{z} + 1/\underline{z}).
\end{aligned} \tag{6.247}$$

Substituting these quantities into Equation 6.245 allows the integral to be converted to one in the complex plane:

$$I = \int_0^{2\pi} d\theta\, w(\sin\theta, \cos\theta) = \oint_C \frac{d\underline{z}}{i\underline{z}}\, w\left(\frac{\underline{z} - 1/\underline{z}}{2i}, \frac{\underline{z} + 1/\underline{z}}{2}\right). \tag{6.248}$$

Example 6.13 As an example of this technique consider the real integral

$$I = \int_0^{2\pi} \frac{d\theta}{2 + \cos\theta}. \tag{6.249}$$

Applying Equations 6.246 and 6.247, this integration becomes

$$I = \int_0^{2\pi} \frac{d\theta}{2 + \cos\theta} = -2i \oint_C \frac{d\underline{z}}{\underline{z}^2 + 4\underline{z} + 1}, \tag{6.250}$$

where C is the counterclockwise contour that lies along the unit circle shown in Figure 6.41. The integrand has two first-order poles at $\underline{z} = -2 \pm \sqrt{3}$. Only the pole at $-2 + \sqrt{3}$ is inside the contour, and it has a residue of $1/(2\sqrt{3})$, so the value of
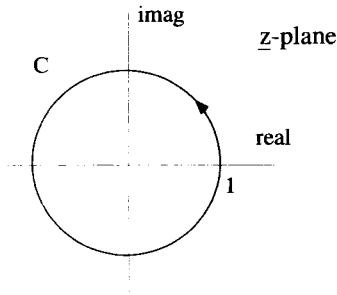
Figure 6.41   Contour for Real sin, cos Integrals

the integral becomes

$$I = \int_0^{2\pi} \frac{d\theta}{2 + \cos\theta}$$

$$= 2\pi i \left(\frac{-2i}{2\sqrt{3}} \right) \tag{6.251}$$

$$= \frac{2\pi}{\sqrt{3}}. \tag{6.252}$$

6.12   CONFORMAL MAPPING

Conformal Mapping is a technique for finding solutions to Laplace's equation in two dimensions. These solutions are of interest in fluid mechanics and electro- and magnetostatics.

The mapping technique uses a complex function, such as $\underline{w} = \underline{w}(\underline{z})$, to take points from the $\underline{z}$-plane and "map" them onto points in the $\underline{w}$-plane. If $\underline{w}_o = \underline{w}(\underline{z}_o)$, then $\underline{w}(\underline{z})$ is said to map the point $\underline{z}_o$ on to the point $\underline{w}_o$, as shown in Figure 6.42. In this figure, we have continued to use the convention that the real and imaginary parts of $\underline{z}$ are given by x and y, and the real and imaginary parts of $\underline{w}$ by u and v. If one point in the
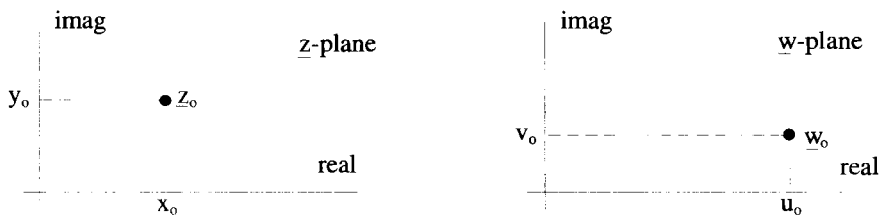
Figure 6.42   A Mapping of $\underline{z}_o$ to $\underline{w}_o$ by $\underline{w} = \underline{w}(\underline{z})$

z-plane maps to only one point in the w-plane, the mapping function $w(z)$ is said to be single valued. It is possible for the mapping function to take one point in the z-plane to more than one point in the w-plane. This type of function is called multivalued.

Mapping is a two-way street. The function $w(z)$ can be inverted to give a function of w, i.e., $z(w)$, that takes points from the w-plane and maps them back to points in the z-plane. If both $w(z)$ and $z(w)$ are single valued, the mapping between the two complex planes is called "one to one."

6.12.1   Mapping of Grid Lines

Many times it is easier to discuss mapping properties with the use of a specific mapping function. For the purposes of this discussion, consider the mapping function

$$w = z^2. \tag{6.253}$$
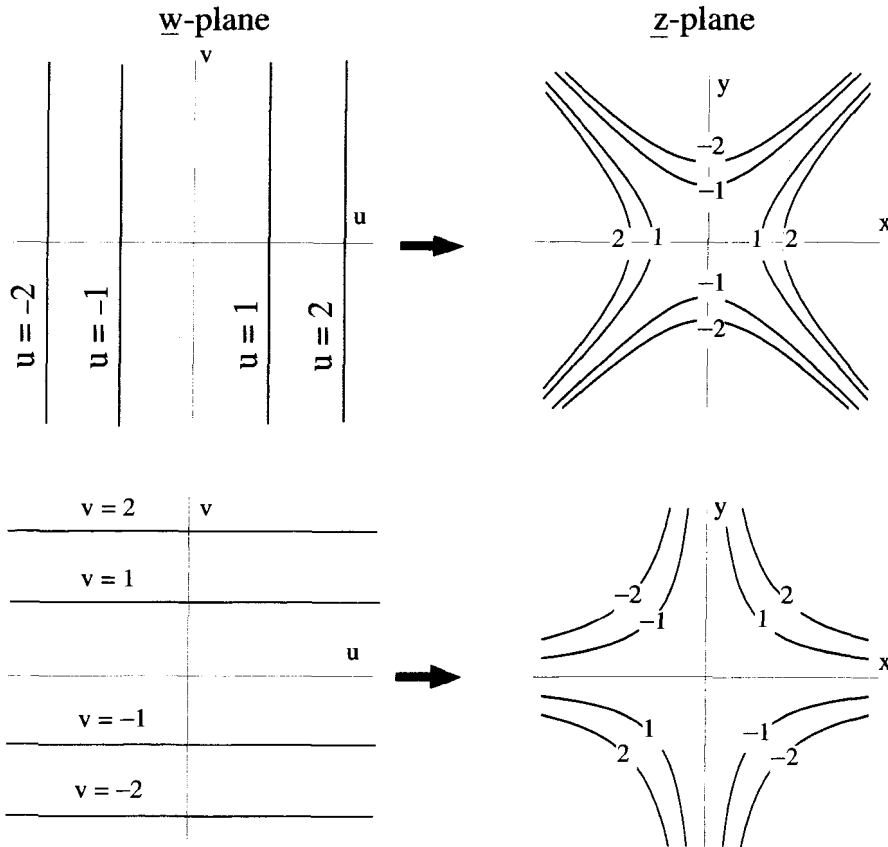
Figure 6.43   Mapping of w-Plane Grid Lines onto z-Plane by $w = z^2$
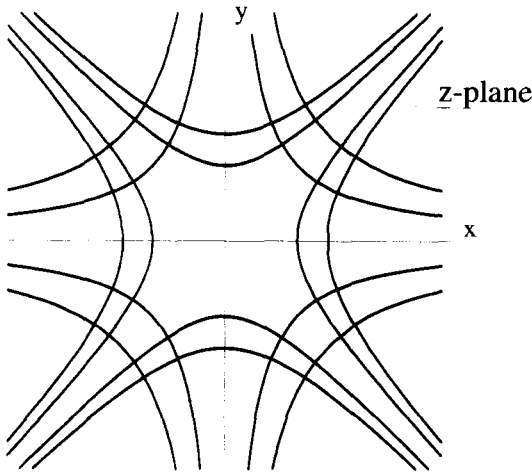
Figure 6.44 Perpendicular Intersection of Grid Line Mapping

How do the grid lines in the w-plane, the lines of constant u and constant v, map onto the z-plane? Substituting $w = u + iv$ and $z = x + iy$ into this mapping function gives

$$u(x, y) = x^2 - y^2$$

$$v(x, y) = 2xy. \qquad (6.254)$$

Equations 6.254 implies that lines of constant u and v represent two sets of nested hyperbola, as shown in Figure 6.43. These hyperbola are plotted on the same graph in Figure 6.44. It can be seen that the hyperbola appear to intersect at right angles. Indeed, we will show that this is the case and is a general property associated with all analytic mapping functions. Notice that the mapping is not one to one, because a single grid line of the w-plane maps into two lines in the z-plane.

6.12.2 Conformal Maps

A mapping function is said to be *conformal* if it preserves angles. This means that a conformal mapping function will map two lines that intersect at some angle into another pair of lines that intersect at the same angle. This is depicted in Figure 6.45,
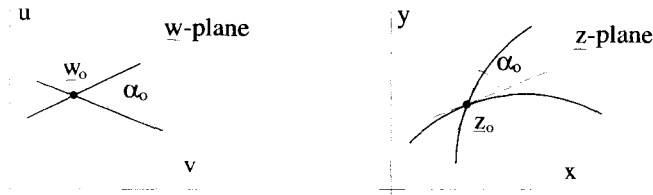
Figure 6.45 The Angle Preserving Feature of Conformal Mapping
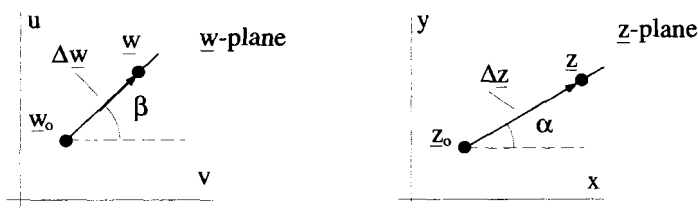
Figure 6.46   Analytic Functions and Conformal Mapping

where the intersection point $\underline{z}_o$ maps into the intersection point $\underline{w}_o$. In both planes, the angle of intersection is α_o. Notice that the lines are not necessarily straight and so the intersection angle, in the case of curved lines, must be defined by the local tangents. The $\underline{w} = \underline{z}^2$ mapping function just discussed appears to be conformal because the perpendicular grid lines of the $\underline{w}$-plane map into hyperbola in the $\underline{z}$-plane that intersect at right angles.

All analytic mapping functions are conformal. To demonstrate this, let a mapping function take the point $\underline{z}_o$ to the point $\underline{w}_o$, and consider a nearby point $\underline{z} = \underline{z}_o + \Delta\underline{z}$ that maps to the point $\underline{w} = \underline{w}_o + \Delta\underline{w}$, in the limit $\Delta\underline{z} \rightarrow 0$. This is shown in Figure 6.46. A line that passes through the two points $\underline{z}_o$ and $\underline{z}$ maps into a line that passes through the points $\underline{w}_o$ and $\underline{w}$. If $\Delta\underline{z} = \Delta z e^{i\alpha}$ and $\Delta\underline{w} = \Delta w e^{i\beta}$, we can write

$$\frac{d\underline{w}}{d\underline{z}} = \lim_{\Delta z \rightarrow 0} \frac{\Delta w}{\Delta z} e^{i(\beta - \alpha)}. \qquad (6.255)$$

But if the mapping function is analytic, we know the derivative at this point does not depend on the way $\Delta\underline{z}$ approaches zero. Thus the right side of Equation 6.255 must be the same for any choice of α in the $\underline{z}$-plane. This means that as α is changed, β must also change in such a way that

$$\beta - \alpha = c_o, \qquad (6.256)$$

where c_o is a constant. All infinitesimal lines passing through z_o will be rotated through the same angle as they are mapped onto the $\underline{w}$-plane. Therefore, all complex functions which are analytic are conformal mapping functions.

6.12.3   Solutions to Laplace's Equation

Conformal mapping functions provide a method for solving Laplace's equation in two dimensions. In a two-dimensional Cartesian geometry, Laplace's equation is

$$\nabla^2 \Phi(x, y) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \Phi(x, y) = 0. \qquad (6.257)$$

If a mapping function $\underline{w} = \underline{w}(\underline{z})$ with

$$\underline{z} = x + iy \qquad (6.258)$$

and

$$\underline{w} = u(x, y) + iv(x, y) \tag{6.259}$$

is analytic, the functions $u(x, y)$ and $v(x, y)$ must satisfy the Cauchy-Riemann conditions

$$\frac{\partial u(x, y)}{\partial x} = \frac{\partial v(x, y)}{\partial y}$$

$$\frac{\partial u(x, y)}{\partial y} = -\frac{\partial v(x, y)}{\partial x}, \tag{6.260}$$

and these partial derivatives must all be continuous. Consequently, the second derivatives obey

$$\frac{\partial^2 u(x, y)}{\partial x^2} = \frac{\partial^2 v(x, y)}{\partial x \partial y}$$

$$\frac{\partial^2 u(x, y)}{\partial y^2} = -\frac{\partial^2 v(x, y)}{\partial y \partial x}. \tag{6.261}$$

This says that

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) u(x, y) = \nabla^2 u(x, y) = 0. \tag{6.262}$$

Equations 6.260 can be manipulated to give the same result for $v(x, y)$:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) v(x, y) = \nabla^2 v(x, y) = 0. \tag{6.263}$$

Therefore, if a mapping function is analytic, the functions $u(x, y)$ and $v(x, y)$ each satisfy a two-dimensional Laplace equation. Either one could be interpreted, for example, as an electrostatic potential or as the pressure in a fluid flow problem.

Example 6.14 In electrostatics, the electric potential satisfies Laplace's equation in a charge-free region. Conformal mapping functions can therefore be used to generate solutions to two-dimensional electrostatic problems. Consider the function

$$\underline{w} = \underline{z}^2. \tag{6.264}$$

As discussed above, the real part of $\underline{w}(\underline{z})$ is a solution to Laplace's equation and can be identified with an electrostatic potential function $\Phi(x, y)$. In this case,

$$\Phi(x, y) = u(x, y) = x^2 - y^2. \tag{6.265}$$

The two-dimensional surfaces of constant potential are given by

$$x^2 - y^2 = \text{Constant}. \tag{6.266}$$

For example, the equipotential $\Phi = 5$ surface obeys

$$x^2 - y^2 = 5, \tag{6.267}$$

and $\Phi = 0$ on the surface

$$x^2 - y^2 = 0. \tag{6.268}$$

This last surface is written more simply as

$$x = \pm y. \tag{6.269}$$

Both these surfaces have been shown in Figure 6.47. This means that if plates were placed on the $x = \pm y$ and $x^2 - y^2 = 5$ surfaces and held at $\Phi = 0$ and $\Phi = 5$ respectively, the electric potential in the space between these plates would be given by

$$\Phi = x^2 - y^2. \tag{6.270}$$

This is shown in Figure 6.48, where we have drawn the equipotential surfaces between the plates for $\Phi = 1, 2, 3$ and 4.

The electric field lines between these plates can also be determined by the mapping function. The relationship between the electric potential and the electric field is

$$\overline{\mathbf{E}} = -\overline{\boldsymbol{\nabla}}\Phi. \tag{6.271}$$

Because $\Phi = u(x, y)$,

$$\overline{\mathbf{E}} = -\frac{\partial u(x, y)}{\partial x}\hat{\mathbf{e}}_x - \frac{\partial u(x, y)}{\partial y}\hat{\mathbf{e}}_y. \tag{6.272}$$

Since field lines always run tangent to the field, if $y = y(x)$ is a field line, then

$$\frac{dy(x)}{dx} = \frac{E_y}{E_x} = \frac{-\partial u(x, y)/\partial y}{-\partial u(x, y)/\partial x}. \tag{6.273}$$
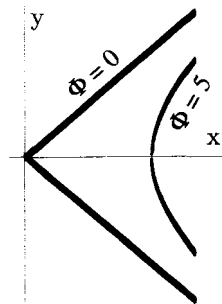
Figure 6.47 Boundary Conditions for the Electrostatic Potential Between Two Sheets
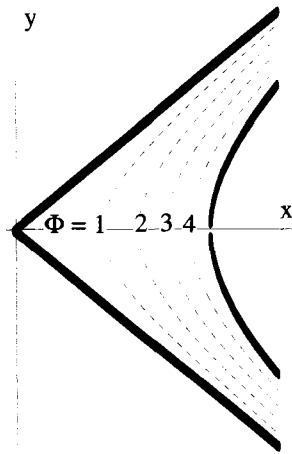
Figure 6.48   Equipotential Lines Between the Two Sheets

But this mapping function is analytic, and so its real and imaginary parts satisfy the Cauchy-Riemann conditions. Consequently,

$$\frac{dy(x)}{dx} = -\frac{\partial v(x, y)/\partial x}{\partial v(x, y)/\partial y}. \tag{6.274}$$

Equation 6.274 can be rearranged to give

$$\frac{\partial v(x, y)}{\partial x}dx + \frac{\partial v(x, y)}{\partial y}dy = 0. \tag{6.275}$$

The left side of Equation 6.275 is just the total differential of $v(x, y)$, so the field lines are simply lines of constant $v(x, y)$. For this particular problem, the field lines are given by $v(x, y) = 2xy$ and are shown in Figure 6.49.

We could also interpret this solution as the laminar flow of a fluid, guided by the surfaces $u = 0$ and $u = 5$. In this case, the lines of constant u would be the flow lines, while the lines of constant v would be lines of constant pressure.

Any conformal mapping function will generate solutions to Laplace's equation. One approach to solving Laplace's equation in various geometries would be to generate a large table of mapping functions, with a corresponding list of boundary geometries, and then, much like using an integral table, find the solution to any given problem by looking it up. This is cumbersome, at best. A more direct way to arrive at the proper mapping function, given the boundary geometry, is presented in the next section.
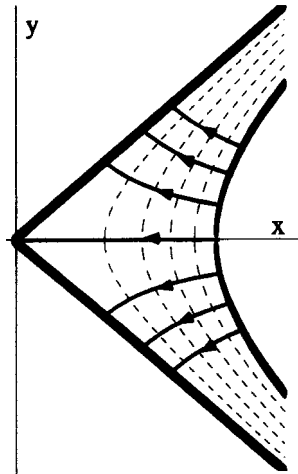
Figure 6.49   Electric Field Lines Between the Two Sheets

6.12.4   Schwartz-Christoffel Mapping Functions

The previously outlined method for obtaining solutions to Laplace's equation using conformal mapping techniques approached the problem from the wrong direction. The mapping function, with solutions to Laplace's equation given by its real and imaginary parts, was specified, and then the boundary conditions associated with the solutions were identified. These types of problems are usually posed the other way around. The Schwartz-Christoffel method uses conformal mapping techniques to design a mapping function from a particular set of boundary conditions.

To see how this method works, begin by considering mapping function $\underline{w} = \underline{w}(\underline{z})$ and its inversion $\underline{z} = \underline{z}(\underline{w})$ such that

$$\frac{d\underline{z}}{d\underline{w}} = \underline{A}(\underline{w} - w_o)^{-k_o}, \tag{6.276}$$

where w_o and k_o are pure real numbers, and $\underline{A}$ is a complex number. We then ask, given the functional form imposed by Equation 6.276, how does the $v = 0$ line map onto the $\underline{z}$-plane? In other words, as $\underline{w}$ moves from $-\infty$ to $+\infty$ along the real $\underline{w}$-axis, as shown in Figure 6.50, what sort of line is traced out in the $\underline{z}$-plane? The quantity
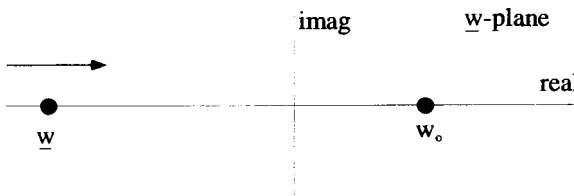
Figure 6.50   Real $\underline{w}$ Mapping
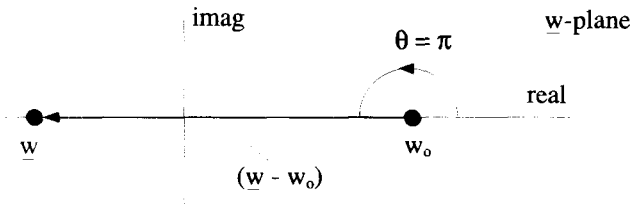
Figure 6.51   Phase of $(\underline{w} - w_o)$ for $\underline{w}$ to the Left of w_o

$d\underline{z}/d\underline{w}$, on the LHS of Equation 6.276, is complex and, in polar notation, has both a magnitude and a phase. Given the form of Equation 6.276, the phase of $d\underline{z}/d\underline{w}$ is

$$\angle \frac{d\underline{z}}{d\underline{w}} = \angle \underline{A} - k_o \angle (\underline{w} - w_o), \tag{6.277}$$

where the operator $\angle$ generates the phase of a complex number. Now consider the phase of $d\underline{z}/d\underline{w}$ as $\underline{w}$ moves from $-\infty$ to $+\infty$ along the real $\underline{w}$-axis. For a pure real $\underline{w}$, with $\underline{w}$ to the left of w_o, Figure 6.51 shows that $\angle (\underline{w} - w_o) = \pi$. For pure real $\underline{w}$, with $\underline{w}$ to the right of w_o, Figure 6.52 shows that $\angle (\underline{w} - w_o) = 0$. Therefore, for $\underline{w}$ on the real axis:

$$\angle \frac{d\underline{z}}{d\underline{w}} = \begin{cases} \angle \underline{A} - k_o \pi & \text{for } \underline{w} \text{ to the left of } w_o \\ \angle \underline{A} & \text{for } \underline{w} \text{ to the right of } w_o \end{cases}. \tag{6.278}$$

As $\underline{w}$ moves along the real $\underline{w}$-axis the mapping function will generate a line in the $\underline{z}$-plane. As $\underline{w}$ is incremented by an amount $\Delta \underline{w}$, $\underline{z}$ will change by an amount $\Delta \underline{z} = (d\underline{z}/d\underline{w})\Delta \underline{w}$. The phases of these $\Delta \underline{w}$ and $\Delta \underline{z}$ steps are related by

$$\angle \Delta \underline{z} = \angle \frac{d\underline{z}}{d\underline{w}} + \angle \Delta \underline{w}. \tag{6.279}$$

As $\underline{w}$ moves to the right along the real axis, $\angle \Delta \underline{w} = 0$ so that

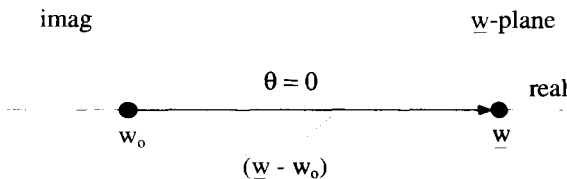$$\angle \Delta \underline{z} = \angle \frac{d\underline{z}}{d\underline{w}}. \tag{6.280}$$

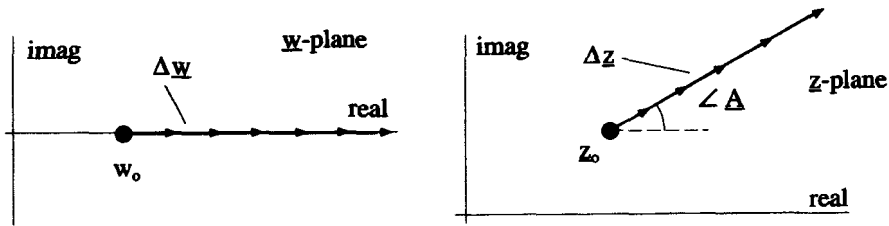**Figure 6.52**   Phase of $(\underline{w} - w_o)$ for $\underline{w}$ to the Right of w_o

Figure 6.53   The Mapping of the Real Axis from w_o to $+\infty$ onto the z-Plane

Let the point w_o map into z_o. Then as w moves away from w_o toward $+\infty$ in steps of Δw, z moves away from z_o in steps of Δz. The result of the mapping of this part of the real w-axis is shown in Figure 6.53. The series of Δw's, all at an angle of zero and all of the same length, are mapped into a series of Δz's. The Δz's are not necessarily the same length but, by Equations 6.278 and 6.280, they all have the same phase, $\angle A$. The mapping of the real w-axis from $-\infty$ to w_o can be done in the same way, but now all the Δz's are at an angle equal to $\angle A$ minus $k_o \pi$, as shown in Figure 6.54. Consequently, the mapping function that obeys Equation 6.276 maps the real w-axis into a line on the z-plane made up of two straight-line segments, as shown in Figure 6.55. There is a break point in this line at z_o, where the slope of the line changes by $k_o \pi$.

These ideas can be extended to a mapping function whose derivative has the form

$$\frac{dz}{dw} = A(w - w_1)^{-k_1}(w - w_2)^{-k_2}(w - w_3)^{-k_3} \cdots (w - w_n)^{-k_n}, \qquad (6.281)$$
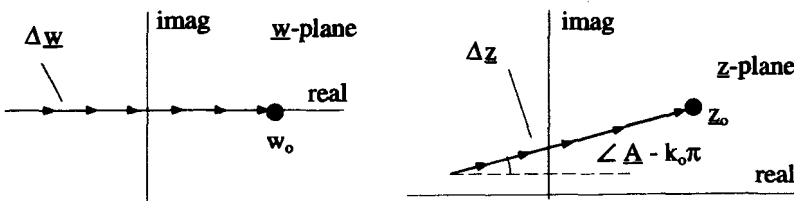
Figure 6.54   The Mapping of the Real Axis from $-\infty$ to w_o onto the z-Plane
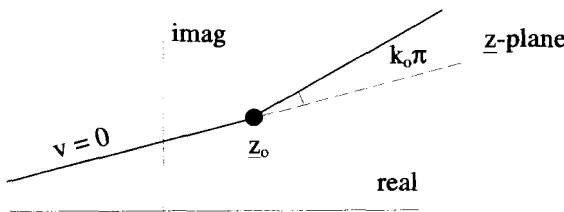
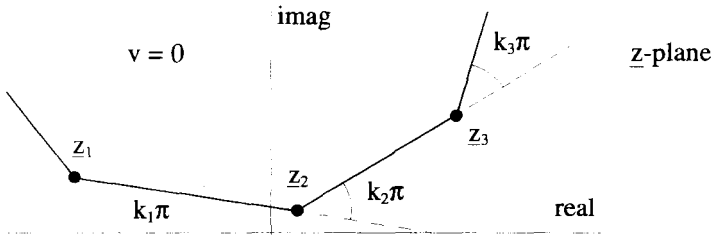**Figure 6.55**   Mapping of the Real w-Axis onto the z-Plane

Figure 6.56   The General Form of a Schwartz-Christoffel Mapping

where w_1, w_2, $w_3 \cdots w_n$ are successively increasing real numbers, and k_1, k_2, k_3, $\cdots k_n$ are arbitrary real numbers. The mapping function that has this form for its derivative maps the real $\underline{w}$-axis into connected straight-line segments in the $\underline{z}$-plane, as shown in Figure 6.56. In this way, the derivative of the mapping function can be set up so that the real $\underline{w}$-axis can be mapped into an arbitrary shape made up of straight-line segments. To obtain the mapping function $\underline{w}(\underline{z})$, Equation 6.281 must be integrated

$$\underline{z}(\underline{w}) = \underline{A} \int d\underline{w} \left\{ (\underline{w} - w_1)^{-k_1} (\underline{w} - w_2)^{-k_2} \right. \tag{6.282}$$

$$\left. (\underline{w} - w_3)^{-k_3} \cdots (\underline{w} - w_n)^{-k_n} \right\}$$

and then inverted.

The real $\underline{w}$-axis can map into a set of straight-line segments that may close to form a polygon in the finite $\underline{z}$-plane, or they may extend to infinity. A variety of boundary shapes can be constructed. Some of the more complicated geometries arise from allowing w_n, the last point along the real $\underline{w}$-axis, to be located at $+\infty$. In this case, the $(\underline{w} - w_n)^{-k_n}$ factor in Equation 6.281 is eliminated, because it can never change the phase of $d\underline{z}/d\underline{w}$. Even though this point does not enter Equation 6.281, it can map into a point $\underline{w}_n$ where there is a $k_n \pi$ angle change. In addition, the Schwartz-Christoffel method can be applied along any horizontal line in the $\underline{w}$-plane. In this way the value of the function on the boundary can assume more than one value.

Example 6.15   In this example, we determine the electric potential and field around a sharp, two-dimensional spike. The spike and ground plane are held at zero volts, while the potential goes to infinity at an infinite distance above the ground plane. This geometry is shown in Figure 6.57. Specifically, $\Phi(x, 0) = 0$ for all x, and also
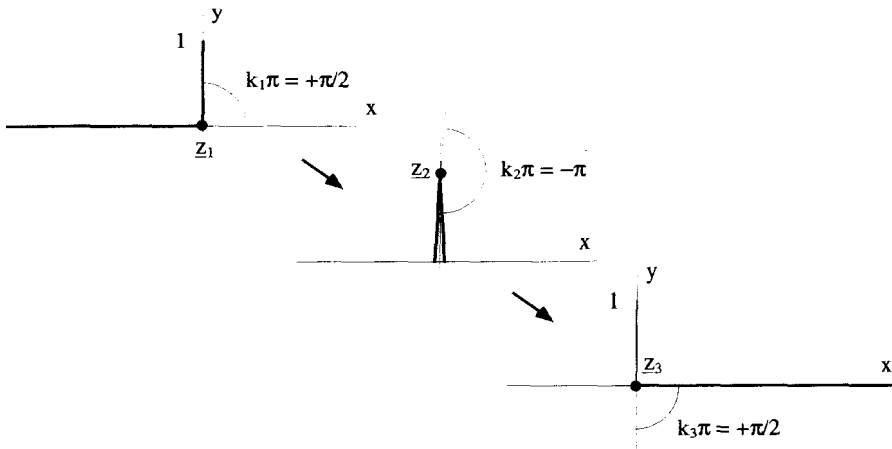

Figure 6.57   Grounded Spike Geometry

Figure 6.58   Mapping of the Real $\underline{w}$-Axis for the Spike Problem

$\Phi(0, y) = 0$ for $0 < y < 1$. We will try to find a solution for Φ in the $y > 0$ half plane.

To use the Schwartz-Christoffel mapping for this problem, the real $\underline{w}$-axis must be mapped onto this complicated shape. The line in the $\underline{z}$-plane can be set up with three break points, as shown in Figure 6.58. These three break points, $\underline{z}_1$, $\underline{z}_2$, and $\underline{z}_3$, are the mapping of the points w_1, w_2, and w_3, which all lie on the real $\underline{w}$-axis, as shown in Figure 6.59. At this point, the exact positions of w_1, w_2 and w_3 are arbitrary, except for the condition that $w_1 < w_2 < w_3$. As $\underline{w}$ moves along the real $\underline{w}$-axis, $\underline{z}$ moves along the path of Figure 6.58, first encountering the break point at $\underline{z}_1$, then the one at $\underline{z}_2$, and finally the one at $\underline{z}_3$. At the $\underline{z}_1$ breakpoint, there is a $\pi/2$ counterclockwise bend, and so $k_1 = 1/2$. At the $\underline{z}_2$ breakpoint, there is a π clockwise bend, and so $k_2 = -1$. Finally, at the $\underline{z}_3$ breakpoint, there is another $\pi/2$ counterclockwise bend, and so $k_3 = 1/2$. The equation for $d\underline{z}/d\underline{w}$ can therefore be written as

$$\frac{d\underline{z}}{d\underline{w}} = \underline{A}(\underline{w} - w_1)^{-1/2}(\underline{w} - w_2)^{+1}(\underline{w} - w_3)^{-1/2}. \tag{6.283}$$

At this point, a bit of trial and error is necessary to figure out the constants $\underline{A}$, w_1, w_2, and w_3. We must select these constants so the mapping function $\underline{z} = \underline{z}(\underline{w})$ has $\underline{z}(w_1) = 0$, $\underline{z}(w_2) = i$, and $\underline{z}(w_3) = 0$. Of course, we cannot check our guess until
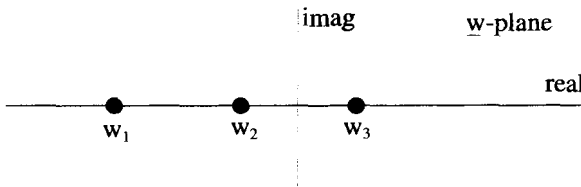

Figure 6.59   Mapping of Break Points onto Real $\underline{w}$-axis

after Equation 6.283 is integrated. From the symmetry in the z-plane, it makes sense to make $w_2 = 0$ and require $w_3 = -w_1$. As a trial set of constants, let $w_1 = -1$, $w_2 = 0$, and $w_3 = 1$ to give

$$\frac{dz}{dw} = A \frac{w}{(w + 1)^{1/2}(w - 1)^{1/2}}. \tag{6.284}$$

The constant A is just a scaling and rotating factor, which can initially be set equal to 1 and changed later if necessary. Equation 6.284 can then be rewritten in the form

$$\frac{dz}{dw} = \frac{w}{\sqrt{w^2 - 1}}, \tag{6.285}$$

which can be reorganized as the indefinite integral

$$\int dz = \int dw \frac{w}{\sqrt{w^2 - 1}}. \tag{6.286}$$

Ignoring the arbitrary constant of integration, the solution of this integral is

$$z = \sqrt{w^2 - 1}. \tag{6.287}$$

Equation 6.287 can be inverted to give

$$w = \sqrt{z^2 + 1}. \tag{6.288}$$

Now this mapping must be checked to see if the constants were chosen properly. When $w = w_1 = -1$, $z = z_1 = 0$. When $w = w_2 = 0$, $z = z_2 = \sqrt{-1} = \pm i$. When $w = w_3 = +1$, $z = z_3 = 0$. So it looks like all the constants are correct. Notice when $w = w_2 = 0$, two values of z_2 are possible. The one that is correct for the geometry of this problem is $z_2 = +i$. The second value simply corresponds to a solution in the lower half of the z-plane.

A check of Equation 6.288 will show that it is analytic everywhere in the z-plane, except at the points $z = \pm i$. Consequently, this mapping function will generate a solution to Laplace's equation everywhere, except at those two isolated points. To obtain the solution, the functions $u(x, y)$ and $v(x, y)$ must be determined. To accomplish this, rewrite Equation 6.288 as

$$(u + iv)^2 = (x + iy)^2 + 1. \tag{6.289}$$

Expand this equation, and separate its real and imaginary parts to get

$$u^2 - v^2 = x^2 - y^2 + 1 \tag{6.290}$$
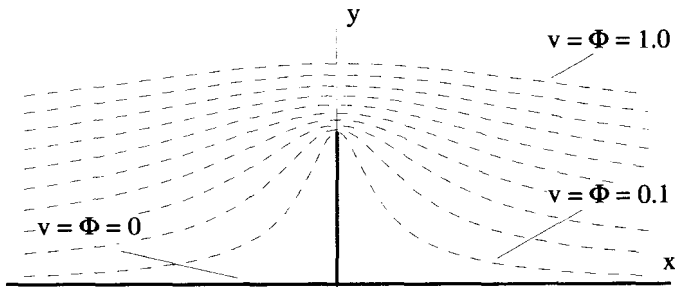
and

$$2uv = 2xy. \tag{6.291}$$

Figure 6.60   Equipotentials above the Grounded Spike

Solving Equation 6.291 for u and substituting the result into Equation 6.290 gives

$$x^2y^2 - v^4 - x^2v^2 + y^2v^2 - v^2 = 0. \qquad (6.292)$$

This equation can be used to generate lines of constant v in the z-plane, which we interpret as lines of constant potential. The $\Phi = v = 0$ line is just the ground plane and spike of Figure 6.57. Several other, equally spaced, equipotential lines obtained from Equation 6.292 are shown in Figure 6.60. A computer is useful in making these plots.

The electric field lines are determined by setting $u(x, y) = $ constant. The equation for these lines is obtained by eliminating v from Equations 6.290 and 6.291:

$$u^4 - x^2y^2 - x^2u^2 + y^2u^2 - u^2 = 0. \qquad (6.293)$$

A plot of these field lines, for uniform steps of u between 0 and 1.4 is shown in Figure 6.61. Remember that in field line drawings such as Figure 6.61, the electric field is not necessarily constant along a field line. The intensity is proportional to the density of the field lines. So in Figure 6.61 it can be seen that the electric field is very strong near the tip of the spike, weakest at the bottom of the spike, and becomes constant far from the spike.
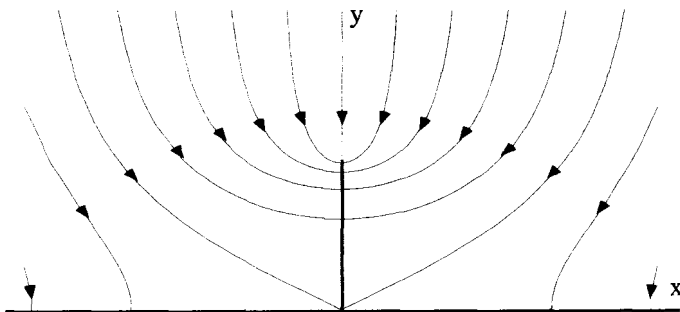

Figure 6.61   Electric Field Lines above the Grounded Spike

EXERCISES FOR CHAPTER 6

1. Use Euler's formula to show that

$$\cos(n\theta) + i\sin(n\theta) = (\cos\theta + i\sin\theta)^n.$$

   This result is called DeMoivre's formula.

2. Find all the zeros of these complex functions:
   i. $\sin z$.
   ii. $\cos z$.
   iii. $\sinh z$.
   iv. $\cosh z$.

3. Make a sketch in the complex z-plane of the following relationships:
   i. $(z - 5)(z^* - 5) = 4$.
   ii. $(z - 5i)(z^* + 5i) = 4$.
   iii. $z^2 + (z^2)^* = 0$.
   iv. $|z| = 2 + \dfrac{z + z^*}{|z|}$.

4. Show that natural logarithms exist for real, negative numbers in the complex plane. In particular, find the value of $\ln(-10)$.

5. Determine all the values of i^i and plot them in the complex plane. As a first step, write i in polar form.

6. Find the real and imaginary parts of the following complex functions:
   i. $w(z) = e^z$.
   ii. $w(z) = \ln z$.
   iii. $w(z) = \sin z$.
   iv. $w(z) = 1/z$.
   v. $w(z) = 1/z^2$.
   vi. $w(z) = \ln(z^2 + 1)$.
   vii. $w(z) = e^{1/z}$.
   viii. $w(z) = 1/\cos z$.
   ix. $w(z) = z/z^*$.
   x. $w(z) = z^* \ln z$.
   In what region of the complex plane are each of these functions analytic?

7. Find an analytic complex function that has a real part of $u(x, y) = x^3 - 3xy^2$.

8. Using the polar forms for z and w,

$$z = r\,e^{i\theta}$$

$$w = R\,e^{i\Theta},$$

a complex functional relationship between $\underline{w}$ and $\underline{z}$ can be written as

$$\underline{w}(\underline{z}) = R(r, \theta)\, e^{i\Theta(r,\theta)}.$$

The functions $R(r, \theta)$ and $\Theta(r, \theta)$ are real functions of the real variables r and θ. These functions are the polar analogues of $u(x, y)$ and $v(x, y)$. If $\underline{w}(\underline{z})$ is analytic, show that the Cauchy Riemann conditions become

$$\frac{\partial R}{\partial r} = \frac{R}{r}\frac{\partial \Theta}{\partial \theta}$$

and

$$\frac{1}{r}\frac{\partial R}{\partial \theta} = -R\frac{\partial \Theta}{\partial r}.$$

9. Show by direct integration that

$$\oint_C d\underline{z}(\underline{z} - \underline{z}_o)^n = \begin{cases} 2\pi i & n = -1 \\ 0 & n \neq -1 \end{cases},$$

where C is any counterclockwise loop that encloses $\underline{z}_o$, and n is an integer.

10. Develop the Taylor series and identify the region of convergence for the following functions:

   i. $\underline{w}(\underline{z}) = \underline{z}$ expanded about $\underline{z} = 2$.

   ii. $\underline{w}(\underline{z}) = \cos \underline{z}$ expanded around $\underline{z} = 0$.

   iii. $\underline{w}(\underline{z}) = \cos \underline{z}$ expanded around $\underline{z} = \pi/2$.

   iv. $\underline{w}(\underline{z}) = \ln(\underline{z} + 1)$ expanded around $\underline{z} = 0$.

   v. $\underline{w}(\underline{z}) = \sqrt{1 - \underline{z}^2}$ expanded about $\underline{z} = 0$.   (up to terms of $\underline{z}^3$)

11. Develop the Laurent series and identify the region of convergence for the following functions:

   i. $\underline{w}(\underline{z}) = e^{1/\underline{z}}$ expanded about $\underline{z} = 0$.

   ii. $\underline{w}(\underline{z}) = 1/(\underline{z} - 1)$ expanded about $\underline{z} = 2$.

   iii. $\underline{w}(\underline{z}) = \dfrac{\underline{z}^2}{(\underline{z} - 2)^2}$ expanded about $\underline{z} = 2$.

   iv. $\underline{w}(\underline{z}) = \dfrac{\underline{z}^2}{(\underline{z} - 2)^2}$ expanded about $\underline{z} = 1$.

12. Find the Taylor and Laurent series expansions around $\underline{z} = 0$ and indicate the regions of convergence for the functions:

   i. $\underline{w}(\underline{z}) = \dfrac{1}{(\underline{z} - 2)^2}.$

   ii. $\underline{w}(\underline{z}) = \dfrac{\cos(\underline{z})}{\underline{z} - 1}.$

iii. $\underline{w}(\underline{z}) = \dfrac{e^{\underline{z}}}{\underline{z}^2 - 1}.$

13. Use the Cauchy integral approach to determine the Taylor series expansion and region of convergence of the following functions around an arbitrary point $\underline{z}_o$:
    (a) $\underline{w}(\underline{z}) = \underline{z}.$
    (b) $\underline{w}(\underline{z}) = \underline{z}^2.$
    (c) $\underline{w}(\underline{z}) = \underline{z} - \underline{a}.$
    (d) $\underline{w}(\underline{z}) = \underline{z}^2/(\underline{z} - \underline{a}).$

14. Using the expansion in Equation 6.100,
    (a) Find the Taylor series expansion about $\underline{z} = 1 + i$ for

    $$\underline{w}(\underline{z}) = \frac{1}{1 + \underline{z}^2}.$$

    (b) Find the Laurent series expansion about the same point for the same function.

15. Using the Cauchy Integration approach, and the complex function

    $$\underline{w}(\underline{z}) = \frac{1}{1 + \underline{z}^2},$$

    (a) Determine the Taylor series expansion for $\underline{w}(\underline{z})$ about $\underline{z} = 0$ and indicate the region of convergence.
    (b) Determine the Laurent series expansion for $\underline{w}(\underline{z})$ about $\underline{z} = 1$ and indicate the region of convergence.

16. Suppose we need to expand the function

    $$\underline{w}(\underline{z}) = \frac{1}{\underline{z} + 1}$$

    with a series that is valid on the closed contour C, a unit circle centered at $\underline{z} = 1$, as shown below:
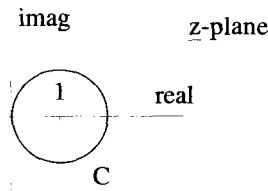
    (a) First consider the Taylor series expansion for $\underline{w}(\underline{z})$ about $\underline{z} = \underline{z}_o$,

    $$\frac{1}{\underline{z} + 1} = \sum_{n=0}^{\infty} \underline{c}_n(\underline{z} - \underline{z}_o)^n.$$

Identify the region of the z-plane where z_o can be placed so that this series will be a valid representation of the function on the contour C and determine the c_n.

(b) Now consider the Laurent series expansion for $w(z)$ about $z = z_o$,

$$\frac{1}{z + 1} = \sum_{n=-\infty}^{\infty} c_n(z - z_o)^n.$$

Identify the region of the z-plane where z_o can be placed so that this series will be a valid representation of the function on the contour C and determine the c_n.

17. Consider the complex function

$$w(z) = \frac{1}{z - 3} + \frac{1}{z - 4}.$$

(a) Find the Taylor series expansion for $w(z)$ about $z = 1$ and identify its region of convergence.

(b) Find the Laurent series expansion for $w(z)$ about $z = 0$, which converges in the annulus $3 < |z| < 4$.

(c) Find the Laurent series expansion for $w(z)$ about $z = 1$, which converges in the annulus $2 < |z - 1| < 3$.

(d) Compare the c_{-1} coefficients in the series solutions to parts (b) and (c) above and comment.

(e) Find the Laurent expansion for $w(z)$ about $z = 1$, which converges in the region $3 < |z - 1|$.

(f) How many different Taylor and Laurent series expansions for $w(z)$ can be made about the point $z = 3.5$?

18. Consider the function

$$w(z) = \frac{e^{2z}}{z - 1}.$$

(a) Find the function that generates the coefficients of the Taylor series expansion for this function about the general point $z = z_o$. Indicate the region of convergence for this series.

(b) Find the function that generates the coefficients of the Laurent series expansion for this function about the general point $z = z_o$. Indicate the region of convergence for this series.

19. Consider the complex function

$$w(z) = \frac{\cos(z)}{z^2}.$$

(a) Find the Laurent series for $\underline{w}(\underline{z})$ expanded about the point $\underline{z} = 1$ and identify the region of convergence.

(b) Find the Taylor series for $\underline{w}(\underline{z})$ expanded about the point $\underline{z} = 1$ and identify the region of convergence.

(c) Find the Laurent series for $\underline{w}(\underline{z})$ expanded about the point $\underline{z} = 0$ and identify the region of convergence.

(d) Why is there no part (d) to this problem?

20. Verify that

$$\sinh \underline{z} = \frac{e^{\underline{z}} - e^{-\underline{z}}}{2}$$

by comparing the Taylor series expansions for $\sinh \underline{z}$ and $e^{\underline{z}}$. Comment on the convergence.

21. Prove that a Laurent expansion of a function about a particular point is unique. That is, show if

$$\underline{w}(\underline{z}) = \sum_{n=-n_o}^{\infty} \underline{a}_n(\underline{z} - \underline{z}_0)^n = \sum_{n=-n_o}^{\infty} \underline{b}_n(\underline{z} - \underline{z}_0)^n,$$

then $\underline{a}_n$ must equal $\underline{b}_n$ for all n. Use the Cauchy integral formula.

22. Consider the function

$$\underline{w} = \frac{\tan \underline{z}}{\underline{z} - \pi/2}.$$

(a) Locate all the poles of $\underline{w}(\underline{z})$ in the $\underline{z}$-plane and identify their order.

(b) Find the Taylor series expansion for $\underline{w}(\underline{z})$ about $\underline{z} = 0$. Identify the first three coefficients of this series and indicate the region of convergence.

(c) How many different Laurent series for $\underline{w}(\underline{z})$ can be constructed about $\underline{z} = 0$?

(d) Find a series expansion around $\underline{z} = 0$, either Taylor or Laurent, that is a valid representation of $\underline{w}(\underline{z})$ at $\underline{z} = \pi$. Indicate the region of convergence for this series.

23. Consider the function

$$\underline{w}(\underline{z}) = \frac{1}{\underline{z}^2 + \underline{z} - 2}.$$

(a) Where in the complex $\underline{z}$-plane are the poles of $\underline{w}(\underline{z})$?

(b) Determine the first three terms for the Taylor series expansion of $\underline{w}(\underline{z})$ about $\underline{z} = 0$.

(c) Identify the region of convergence for the Taylor series of part (b).

(d) Determine the general expression for the nth coefficient of the Taylor series expansion of part (b).

(e) There is a Laurent series expansion for $\underline{w}(\underline{z})$ about $\underline{z} = 0$ in the region $1 < |\underline{z}| < 2$ that has the form

$$\underline{w}(\underline{z}) = \sum_{n=-\infty}^{\infty} \underline{c}_n \underline{z}^n.$$

Obtain the general expression for the $\underline{c}_n$.

24. Consider the function

$$\underline{w}(\underline{z}) = \frac{1}{\underline{z}^2 - \underline{z}}.$$

(a) Can this function be expanded in a Taylor series about $\underline{z} = 0$?

(b) Find two different Laurent series expansions for this function, both expanded about $\underline{z} = 0$. Indicate their regions of convergence.

(c) Find the Taylor series for this function expanded about $\underline{z} = 1/2$. Indicate the region of convergence.

(d) Find the Laurent series for this function expanded about $\underline{z} = 1/2$. Indicate the region of convergence.

(e) What is the residue of this function at $\underline{z} = 0$? at $\underline{z} = 1$? at $\underline{z} = 1/2$?

(f) What is the value of the integral

$$\oint_C d\underline{z} \, \underline{w}(\underline{z}),$$

where C is a counterclockwise circle centered at the origin of radius $1/2$? of radius 1.5?

(g) What is the value of the integral in part (f) above if C is a clockwise circle centered at $\underline{z} = 1$ of radius 0.5? of radius 1.5?

25. Identify the locations and orders of the poles for the following functions. For each of the functions evaluate the residue associated with each of its poles.

i. $\underline{w}(\underline{z}) = \dfrac{\underline{z}^2}{\underline{z}^2 + 1}.$

ii. $\underline{w}(\underline{z}) = \dfrac{\sin \underline{z}}{\underline{z}^2(\underline{z} - \pi/2)}.$

iii. $\underline{w}(\underline{z}) = \dfrac{\underline{z}}{\sin \underline{z}}.$

iv. $\underline{w}(\underline{z}) = \dfrac{\ln \underline{z}}{\underline{z}^2 - 1}.$

v. $\underline{w}(\underline{z}) = \dfrac{1}{\underline{z}^2(e^{\underline{z}} - 1)}.$

vi. $\underline{w}(\underline{z}) = \dfrac{\underline{z}^2\, e^{\underline{z}}}{1 + e^{2\underline{z}}}.$

26. Each of the following functions has a pole at $\underline{z} = 1$. In each case identify the order of the pole and determine its residue.

   i. $\dfrac{1}{\underline{z}^2 - 2\underline{z} + 1}.$

   ii. $\dfrac{1}{\underline{z}^2 - 1}.$

   iii. $\dfrac{1}{\ln \underline{z}}.$

   iv. $\dfrac{1 - \cos(\underline{z} - 1)}{(\underline{z} - 1)^3}.$

   v. $\cot(\underline{z} - 1).$

27. Let C be a counterclockwise, circular contour of radius 2 centered at $\underline{z} = 0$. Evaluate the complex integrals:

   i. $\displaystyle\oint_C d\underline{z}\, \dfrac{\cos \underline{z}}{\underline{z}^3 - 1}.$

   ii. $\displaystyle\oint_C d\underline{z}\, \dfrac{\cos \underline{z}}{(\underline{z} - 1)^3}.$

   iii. $\displaystyle\oint_C d\underline{z}\, \dfrac{\sin \underline{z}}{\underline{z}^2(\underline{z} - \pi/2)}.$

   iv. $\displaystyle\oint_C d\underline{z}\, \dfrac{\sin \underline{z}}{\underline{z}^n}$ (n = an integer).

28. Evaluate the integral

$$\oint_C d\underline{z}\, \frac{\underline{z}^2}{(\underline{z} - 1)(\underline{z} - 3)},$$

   (a) where C is the clockwise, circular contour given by $|\underline{z}| = 2$.

   (b) where C is the clockwise, circular contour given by $|\underline{z}| = 4$.

29. Evaluate the complex integral

$$I = \oint_C d\underline{z}\, \frac{\sin \underline{z}}{(\underline{z} - \pi/2)^2(\underline{z} + \pi/2)},$$

where C is the counterclockwise, circular contour given by $|\underline{z} + \pi/2| = 1$.

30. Evaluate the complex integral

$$\oint_C d\underline{z}\, \frac{1}{\underline{z}^3 - 8},$$

where C is the counterclockwise, circular contour $|\underline{z} - 1| = 2$.

31. Evaluate the complex integral

$$\oint_C dz \frac{1}{z^4 + 4},$$

where C is the clockwise, circular contour $|z - 2i| = 1$.

32. Evaluate the complex integral

$$\oint_C dz \frac{\sin z}{z^2 - z/2},$$

where C is the counterclockwise, circular contour $|z| = 1$.

33. Consider the complex integral

$$\oint_C dz \frac{\sin z}{(z - \pi/2)(z + \pi/2)}$$

and the two contours $C = C_1$ and $C = C_2$ shown below.
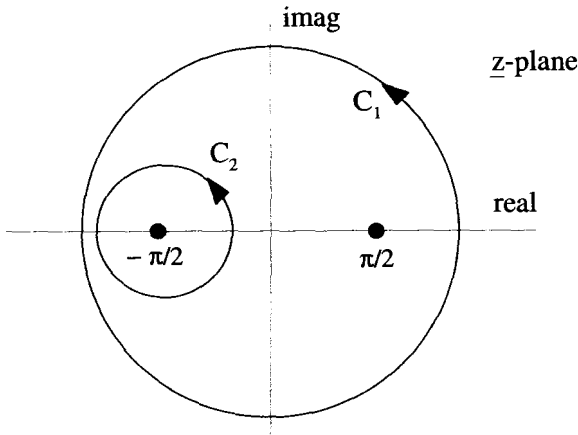
Let I_1 be the integral around C_1 and I_2 the integral around C_2. Show by contour deformation that $I_2 - I_1 = -2i$.

34. Consider the function

$$w(z) = \frac{z}{1 - \sin(2z)}.$$

(a) Find all the poles of $w(z)$ and identify their order.

(b) Evaluate the integral

$$\oint_C dz\, w(z),$$

where C is a counterclockwise, circular contour given by $|z| = 1$.

(c) Evaluate the integral of part (b) above if C is a counterclockwise, circular contour given by $|z - \pi| = 1$.

35. Consider the function

$$w(z) = \frac{z}{1 - \cos z}.$$

(a) Find all the poles of $w(z)$ and identify their order.

(b) Evaluate the integral

$$\oint_C dz\, w(z),$$

where C is a counterclockwise, circular contour given by $|z| = 1$.

(c) Evaluate the integral of part (b) above if C is a counterclockwise, circular contour given by $|z - 2\pi| = 1$.

36. Consider the integral

$$\oint_C dz\, \frac{z}{e^z - 1}.$$

(a) Locate the poles of the integrand and identify their order.

(b) Evaluate the integral if C is a counterclockwise, circle of radius 1 centered at

    i. $z = 0$.

    ii. $z = 2\pi$.

    iii. $z = 2\pi i$.

37. Consider the integral

$$I = \oint_C dz\, e^{1/z} z^{n-1},$$

where n is a positive integer, and C is the counterclockwise circle $|z| = 1$.

(a) Find the Laurent series expansion about $z = 0$ for the integrand.

(b) Use the residue theorem to evaluate the integral.

38. Convert each of the following integrals to an equivalent one in the complex plane. Identify all the poles of the complex integrands and their residues. Close the contours and evaluate the integrals.

    i. $\displaystyle\int_{-\infty}^{\infty} dx\, \frac{\cos x}{x^2 + 1}$.

    ii. $\displaystyle\int_0^{\infty} dx\, \frac{x \sin x}{x^2 + 1}$.

iii. $\displaystyle\int_0^\infty dx \, \frac{\sin x}{x}$.

iv. $\displaystyle\int_0^\infty dx \, \frac{x^2 \cos(2x)}{x^2 + 1}$.

v. $\displaystyle\int_{-\infty}^\infty dx \, \frac{e^x}{\cosh(\pi x)}$.

vi. $\displaystyle\int_0^\infty dx \, \frac{\sinh(ax)}{\sinh(\pi x)} \qquad -\pi < a < \pi$.

39. Convert each of the following real integrals into a complex integral and evaluate it using the residue theorem:

i. $\displaystyle\int_0^\pi d\theta \, \frac{1}{1 + \sin^2 \theta}$.

ii. $\displaystyle\int_0^\pi d\theta \, \frac{1}{1 + \cos^2 \theta}$.

iii. $\displaystyle\int_0^{2\pi} d\theta \, \frac{\cos \theta}{5 + 4 \cos \theta}$.

iv. $\displaystyle\int_0^{2\pi} d\theta \, \frac{\cos(3\theta)}{5 - 4 \cos \theta}$.

40. Show that the integral

$$\int_0^\infty dx \, \frac{1}{1 + x^3},$$

where x is a pure real variable, can be evaluated by doing an integration in the $\underline{z}$-plane around the closed contour shown in the figure below.
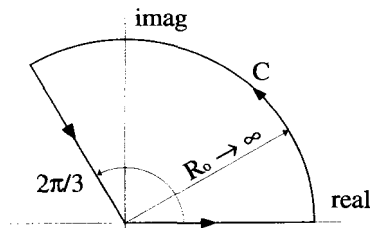
Identify the similar closed contour in the $\underline{z}$-plane that can be used to evaluate the integral

$$\int_0^\infty dx \, \frac{1}{1 + x^m},$$

where m is any positive integer.

41. Consider the integral

$$PP \int_{-\infty}^{\infty} dx \, \frac{1}{(x^2 + 1)(x + 1)},$$

where we must use the principal part because of the behavior of the integrand at $x = -1$. Convert this into an integral in the complex plane and determine an appropriate contour of integration to evaluate this integral. Show that you obtain the same answer regardless of which side of the $\underline{z} = -1$ pole you go around.

42. Consider the integral

$$I(u) = \int_{-\infty}^{\infty} dx \, \frac{e^{-ixu}}{x^2 + 1},$$

where x and u are real variables.

(a) Without performing the integration, show that $I(u)$ is a pure real function of u.

(b) Convert this integral into the complex plane and evaluate it using closure and the residue theorem.

(c) Plot $I(u)$ vs. u.

(d) The function $I(u)$ can be converted to a complex function of a complex variable by substituting the complex variable $\underline{w}$ for u, i.e., $\underline{I}(u) \to I(\underline{w})$. Is $\underline{I}(\underline{w})$ analytic?

43. Use complex integral techniques to evaluate the integral

$$\int_{-\infty}^{\infty} dk \, \frac{\sin(kx)}{k},$$

where k and x are both real variables.

44. Discuss the problems with trying to close the following integral in the complex plane:

$$\int_{-\infty}^{\infty} dx \, \frac{e^{-x^2}}{x^2 + x_o^2}.$$

45. Consider the semicircular contour shown below in the limit as $\epsilon \to 0$.
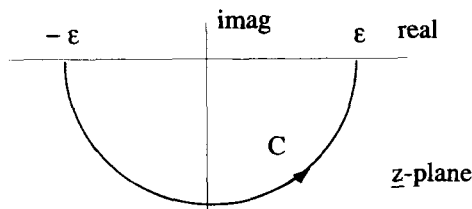
On this contour, evaluate

i. $\displaystyle \int_C d\underline{z}\,\frac{1}{\underline{z}}.$

ii. $\displaystyle \int_C d\underline{z}\,\frac{1}{\underline{z}-i}.$

iii. $\displaystyle \int_C d\underline{z}\,\frac{1}{\underline{z}(\underline{z}-i)}.$

46. Use the results of the previous problem to evaluate

$$\oint d\underline{z}\,\frac{1}{\underline{z}(\underline{z}-i)},$$

where C is the closed contour shown below in the limit $\epsilon \to 0$ and $R_o \to \infty$.
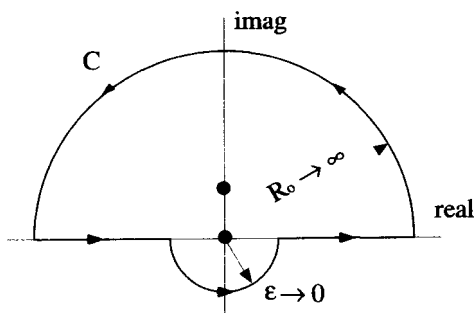
Based on this result, what is the principal part of the integral

$$\text{PP}\int_{-\infty}^{\infty} dx\,\frac{1}{x(x-i)}$$

where x is a real variable?

47. Consider the function

$$\underline{w}(\underline{z}) = \frac{\cos\underline{z}}{\underline{z}^2}.$$

(a) Evaluate the integral

$$\oint_C d\underline{z}\,\underline{w}(\underline{z}),$$

where C is the counterclockwise, circular contour given by $|\underline{z}| = 1$.

(b) Evaluate the principal part of the integral

$$\text{PP}\int_{-\infty}^{\infty} dx\,\frac{\cos x}{x^2}.$$

48. How does the function $\underline{w} = \underline{z}^2$ map the unit circle centered at the origin of the $\underline{z}$-plane onto the $\underline{w}$-plane?

49. Using the mapping function

$$\underline{w}(\underline{z}) = \frac{1}{\underline{z}},$$

  (a) how does the unit circle $|\underline{w} - 1| = 1$ map onto the $\underline{z}$-plane?

  (b) how does the line $y = 1/2$ map onto the $\underline{z}$-plane?

50. For each of the mapping functions below, indicate how the lines of constant u and v in the $\underline{w}$-plane get mapped onto the $\underline{z}$-plane:

  i. $\underline{w}(\underline{z}) = e^{\underline{z}}$.

  ii. $\underline{w}(\underline{z}) = \dfrac{\underline{z} - 1}{\underline{z}}$.

  iii. $\underline{w}(\underline{z}) = \ln \underline{z} - 1$.

51. What part of the $\underline{z}$-plane corresponds to the interior of the unit circle centered at the origin of the $\underline{w}$-plane for the following mapping functions?

  i. $\underline{w}(\underline{z}) = \dfrac{\underline{z} - 1}{\underline{z} + 1}$.

  ii. $\underline{w}(\underline{z}) = \dfrac{\underline{z} - i}{\underline{z} + i}$.

52. Consider the mapping function

$$\underline{w} = \underline{z}\,\underline{z}^* + \underline{z}.$$

  (a) Is this mapping function analytic?

  (b) Show how this function maps the grid lines of the $\underline{z}$-plane onto the $\underline{w}$-plane. Specifically map, on a labeled drawing, the grid lines shown below.
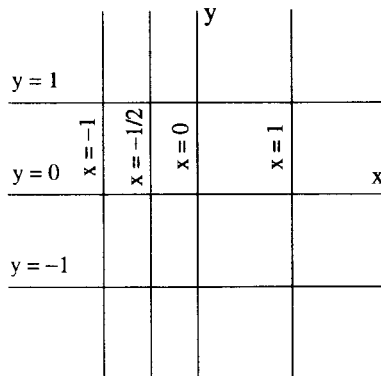
(c) Is this mapping function conformal? Can it be used to generate solutions to Laplace's equation?

53. Consider the mapping function

$$\underline{w}(\underline{z}) = \frac{\underline{z} - 1}{\underline{z} + 1}.$$

(a) Where does the imaginary axis of the $\underline{w}$-plane map onto the $\underline{z}$-plane?

(b) Where do the points in the right half of the $\underline{w}$-plane map onto the $\underline{z}$-plane?

(c) Where do the points in the left half of the $\underline{w}$-plane map onto the $\underline{z}$-plane?

54. The mapping function

$$\underline{w}(\underline{z}) = \frac{\underline{z}^2 + 1}{\underline{z}}$$

can be used to analyze two-dimensional flow around a cylindrical obstacle. Uniform stream lines described in the $\underline{w}$-plane by lines of constant v become, upon mapping onto the $\underline{z}$-plane, the flow lines around a cylinder. The constant pressure lines for the flow around the cylinder are given by mapping the lines of constant u onto the $\underline{z}$-plane.

(a) Plot the lines $v = -3, -2, -1, 0, +1, +2,$ and $+3$ onto the $\underline{z}$-plane.

(b) Plot the lines of constant u onto the $\underline{z}$-plane.

(c) The velocity field for the flow is proportional to the gradient of the pressure distribution function, i.e.,

$$\overline{\mathbf{v}} = -\mu \overline{\nabla} P(x, y),$$

where μ is a positive constant called the mobility, and $P(x, y)$ is the pressure distribution function. What is the velocity field, $\overline{\mathbf{v}}(x, y)$, for the flow around the cylinder?

55. Use the Schwartz-Christoffel method to determine the functional relationship of $d\underline{z}/d\underline{w}$ that takes the real $\underline{w}$-axis to the step in the $\underline{z}$ plane shown below. Integrate and invert this expression to obtain the mapping function $\underline{w} = \underline{w}(\underline{z})$. Make an attempt to plot the lines of constant v onto the $\underline{z}$-plane. If you have access to computer plotting software, you should be able to make very accurate maps of the constant v lines.
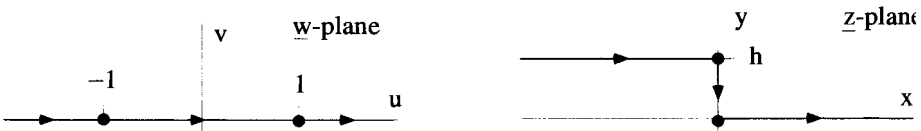
56. Use the Schwartz-Christoffel conformal mapping technique to find the stream function for flow over a surface that is initially flat and then rises at $45°$, as shown below.
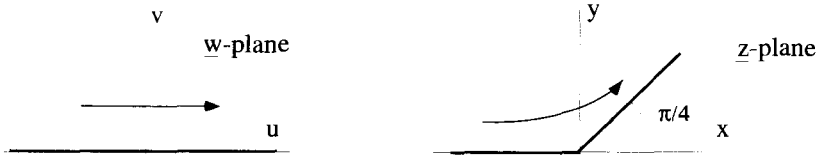
(a) Find the mapping function that takes the real $\underline{w}$-axis into the line in the $\underline{z}$-plane made up of the negative real axis and a segment out to infinity at $\theta = \pi/4$.

(b) Find the function that takes stream lines in the $\underline{w}$-plane to stream lines in the $\underline{z}$-plane. **NOTE:** The last couple of lines of algebra in this calculation get a bit messy, but the answer is obtainable.

57. Recall, the mapping function generated by the Schwartz-Christoffel method is given by the expression

$$\frac{d\underline{z}}{d\underline{w}} = \underline{A}(\underline{w} - w_1)^{-k_1}(\underline{w} - w_2)^{-k_2}(\underline{w} - w_3)^{-k_3} \cdots.$$

(a) How does the real $\underline{w}$-axis map onto the $\underline{z}$-plane if: $\underline{A} = e^{i\pi/4}$, $w_1 = -1$, $k_1 = 1/2$, $w_2 = 0$, $k_2 = -1$, $w_3 = 1$, and $k_3 = 1/2$? Draw a sketch of this mapping.

(b) How would your answer to part (a) change if: $\underline{A} = 1$, $w_1 = -1$, $w_2 = 1/2$, and $w_3 = 2$, while all the k values stayed the same?

58. Using the Schwartz-Christoffel method, design a mapping function that takes the real $\underline{w}$-axis to a combination of the negative real $\underline{z}$-axis and the positive imaginary $\underline{z}$-axis with $\underline{w} = 0$ mapping into $\underline{z} = 0$.
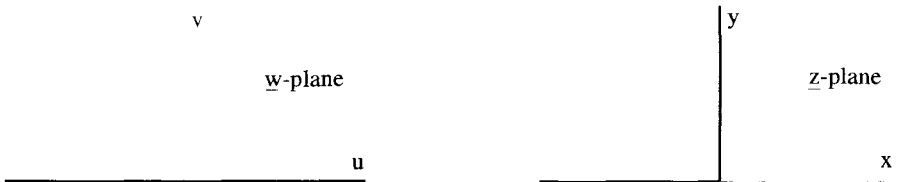
Using this mapping function, map the $45°$ lines shown below in the $\underline{z}$-plane onto the $\underline{w}$-plane.
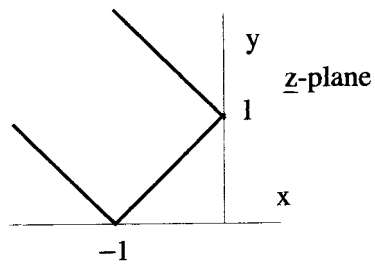
59. Show that if

$$\underline{z} = \sin \underline{w},$$

then

$$\underline{w} = \sin^{-1} \underline{z} = -i \ln \left[i\underline{z} + (1 - \underline{z}^2)^{1/2} \right].$$

7

FOURIER SERIES

This chapter consists of four sections. In the first section, we present a quick refresher of Fourier series methods and a few examples. The second section demonstrates the utility of the complex, exponential form of the Fourier series. These two sections should provide a framework for the subsequent chapters on Fourier and Laplace transforms. The third section briefly discusses the convergence properties of Fourier series and introduces the concept of completeness. In the final section, we discuss the discrete Fourier series, which is a numerical technique for obtaining Fourier spectra.

7.1 THE SINE-COSINE SERIES

A general Fourier series expansion is the sum of a potentially infinite number of sine and cosine terms:

$$s(t) = \frac{a_o}{2} + \sum_{n=1,2,...}^{\infty} a_n \cos \omega_n t + \sum_{n=1,2,...}^{\infty} b_n \sin \omega_n t. \tag{7.1}$$

Each term oscillates at an frequency $\omega_n = 2\pi n/T_o$, an integer multiple of the fundamental frequency $\omega_o = 2\pi/T_o$. The value of T_o is determined either by the periodicity of the function the series is supposed to represent, or a periodicity we force the series expansion to take. The form of Equation 7.1 guarantees that the function generated by the series has a periodicity of T_o. In other words, it obeys the condition

$$s(t) = s(t + T_o) \qquad \text{for all } t. \tag{7.2}$$

An example of such a function is shown in Figure 7.1. If the independent variable is time, then $s(t)$ has gone on for all past time and must continue forever.
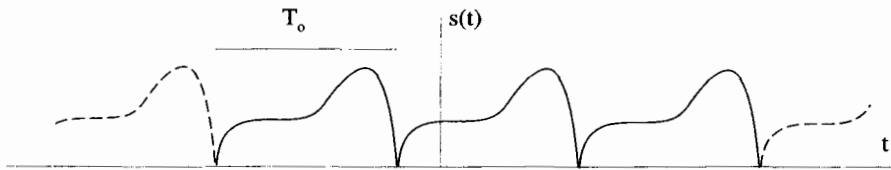
Figure 7.1   Periodic Function with Basic Period T_o

In a typical Fourier series problem, we are given a particular function $f(t)$, and we want to determine both T_o and the a_n and b_n coefficients, so the series is equal to $f(t)$, for all t. For this to be possible, $f(t)$ itself must be periodic and exist for all t, i.e.,

$$f(t) = f(t + T_o) \qquad \text{for all } t. \tag{7.3}$$

The smallest value of the constant T_o that satisfies Equation 7.3 is defined as the *period* of $f(t)$.

If $f(t)$ is not periodic, a Fourier series cannot be used to represent $f(t)$ for all t. We can, however, make the series equal to $f(t)$ over some finite interval. Consider the nonperiodic $f(t)$ shown in Figure 7.2, and the interval $t_1 < t < t_2$. We can define the basic period for the Fourier series to be $T_o = t_2 - t_1$, and the generated series can be made identical to $f(t)$ over this interval, as shown in Figure 7.3. Outside the interval, the Fourier series $s(t)$ is periodic and will not match the function $f(t)$.
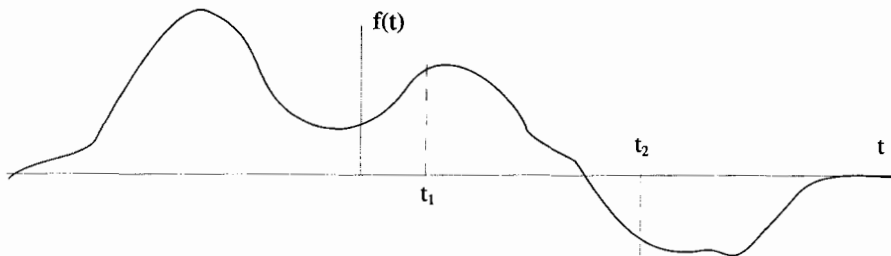
Figure 7.2   A Nonperiodic Function
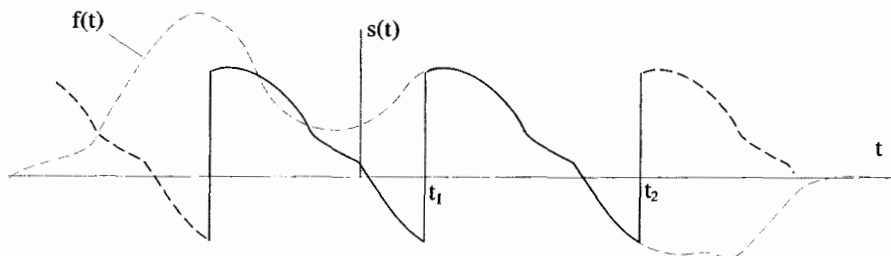
Figure 7.3   Fourier Series Fit to a Nonperiodic Function

7.1.1 The Orthogonality Conditions

To determine the coefficients a_n and b_n, we must first derive a set of equations called the orthogonality conditions. These relations involve integrals of products between the sine and cosine functions, i.e., (sin)(cos), (sin)(sin), and (cos)(cos).

The derivation starts by considering what happens when either the sine or cosine functions are integrated over an integer multiple of a period. For any integer $n \neq 0$, it is obvious that

$$\int_{t_o}^{t_o+T_o} dt \, \sin\left(\frac{2\pi n}{T_o}t\right) = \int_{t_o}^{t_o+T_o} dt \, \cos\left(\frac{2\pi n}{T_o}t\right) = 0. \tag{7.4}$$

The integrals in Equation 7.4, and the ones that follow, are valid for any t_o. For $n = 0$ the first of the above integrals is still zero. However, the $n = 0$ integral for the cosine function takes on the special value

$$\int_{t_o}^{t_o+T_o} dt \, \cos(0) = T_o. \tag{7.5}$$

The first orthogonality condition involves an integral of the product of sine and cosine functions. Because

$$\sin \alpha \cos \beta = \frac{\sin(\alpha + \beta) + \sin(\alpha - \beta)}{2}, \tag{7.6}$$

Equation 7.4 implies

$$\int_{t_o}^{t_o+T_o} dt \, \sin\left(\frac{2\pi n}{T_o}t\right) \cos\left(\frac{2\pi m}{T_o}t\right) = 0 \tag{7.7}$$

for all integer values of n and m including zero. The other orthogonality conditions, for integrals involving the products of the sine function with other sine functions and the cosine function with other cosine functions, are obtained using

$$\sin \alpha \sin \beta = \frac{\cos(\alpha - \beta) - \cos(\alpha + \beta)}{2}$$

$$\cos \alpha \cos \beta = \frac{\cos(\alpha - \beta) + \cos(\alpha + \beta)}{2}. \tag{7.8}$$

With Equations 7.7 and Equation 7.5, we can write for $n > 0$ and $m > 0$

$$\int_{t_o}^{t_o+T_o} dt \, \sin\left(\frac{2\pi n}{T_o}t\right) \sin\left(\frac{2\pi m}{T_o}t\right) = \delta_{mn}\frac{T_o}{2} \tag{7.9}$$

and

$$\int_{t_o}^{t_o+T_o} dt \, \cos\left(\frac{2\pi n}{T_o}t\right) \cos\left(\frac{2\pi m}{T_o}t\right) = \delta_{nm}\frac{T_o}{2}. \tag{7.10}$$

7.1.2 Evaluating the Coefficients

The orthogonality conditions allow us to isolate and evaluate the individual coefficients of the Fourier series. If we want a series to equal $f(t)$, we write

$$f(t) = \frac{a_o}{2} + \sum_{n=1,2,\ldots}^{\infty} a_n \cos \omega_n t + \sum_{n=1,2,\ldots}^{\infty} b_n \sin \omega_n t. \tag{7.11}$$

The a_o coefficient is obtained from

$$\int_{t_o}^{t_o+T_o} dt \, f(t) = \int_{t_o}^{t_o+T_o} dt \left[\frac{a_o}{2} + \sum_{n=1,2,\ldots}^{\infty} a_n \cos \omega_n t + \sum_{n=1,2,\ldots}^{\infty} b_n \sin \omega_n t \right]. \tag{7.12}$$

Using Equation 7.4, we can see that the only term in the expansion that does not integrate to zero is the $a_o/2$ term. Thus

$$\int_{t_o}^{t_o+T_o} dt \, f(t) = \int_{t_o}^{t_o+T_o} dt \, \frac{a_o}{2} = \frac{a_o T_o}{2}, \tag{7.13}$$

or

$$a_o = \frac{2}{T_o} \int_{t_o}^{t_o+T_o} dt \, f(t). \tag{7.14}$$

The remaining a_n coefficients can be obtained in a similar manner. Operating on Equation 7.11 by an integral weighted by $\cos \omega_m t$, where $m > 0$, and using the orthogonality conditions gives

$$\int_{t_o}^{t_o+T_o} dt \, \cos(\omega_m t) f(t) = \sum_{n=1,2,\ldots}^{\infty} a_n \delta_{nm} \frac{T_o}{2} = a_m \frac{T_o}{2}. \tag{7.15}$$

Thus, the a_m^{th} coefficient can be evaluated using the expression

$$a_m = \frac{2}{T_o} \int_{t_o}^{t_o+T_o} dt \, \cos(\omega_m t) f(t). \tag{7.16}$$

Notice that Equation 7.16 is also valid for $m = 0$. This was the reason the factor of 2 was used in the a_o term of Equation 7.11. With that convention, the a_o term does not need to be treated differently from the other a_m terms. The b_n coefficients are obtained in exactly the same manner, except the integrals are performed with $\sin \omega_m t$. The orthogonality equations give

$$\int_{t_o}^{t_o+T_o} dt \, \sin(\omega_m t) f(t) = \sum_{n=1,2,\ldots}^{\infty} b_n \delta_{nm} \frac{T_o}{2} = b_m \frac{T_o}{2} \tag{7.17}$$

with the result that

$$b_m = \frac{2}{T_o} \int_{t_o}^{t_o+T_o} dt \, \sin(\omega_m t) f(t). \tag{7.18}$$

7.1.3 The Fourier Series Circuit

There is an interesting way to view the summation of terms in a Fourier series which we will borrow from the field of electronics. We can model each term in the summation by a signal generator, which puts out a voltage signal at the proper frequency and amplitude for that term. This is shown in Figure 7.4. The a_o term is
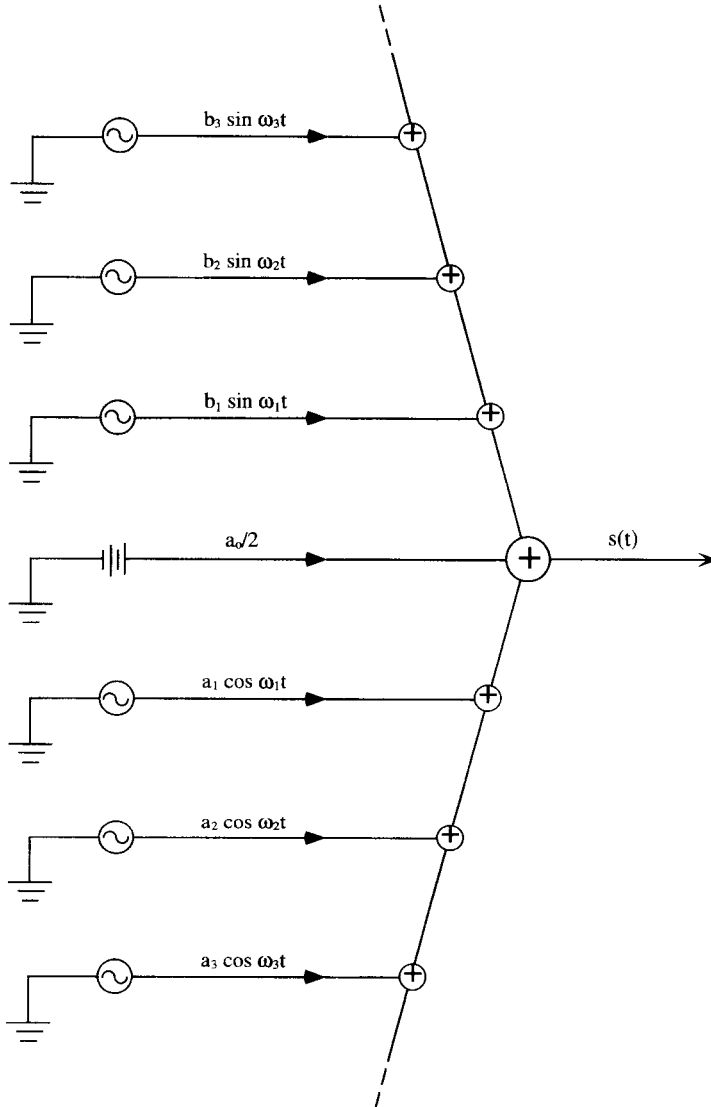
Figure 7.4   Fourier Sine-Cosine Series Circuit

a d.c. source of $a_o/2$ volts. The b_1 term is a sine wave oscillating at frequency ω_1, with a signal amplitude of b_1 volts. The summation is accomplished by joining all the signals into one output from the circuit. This is the basis for what the electrical engineers call the frequency spectrum of a signal. Conceptually, they often break a signal into its various Fourier series components, analyze how each component behaves individually, and then join the results back together to generate their final solution.

We will return to this analogy in more detail when we discuss Fourier and Laplace transforms.

Example 7.1   As a very simple first example, consider the "expansion" of the function, $f(t) = \sin t$. The function is obviously periodic, with $T_o = 2\pi$. The Fourier series takes the form

$$\sin t = \frac{a_o}{2} + \sum_{n=1}^{\infty} a_n \cos(nt) + \sum_{n=1}^{\infty} b_n \sin(nt). \tag{7.19}$$

We could struggle through the integral equations of the previous section to pick out the values of the a_n and b_n coefficients. In this particular case, however, it is easy to see that the only nonzero coefficient is b_1, which has a value of 1.

The set of coefficients in the Fourier series is sometimes referred to as the *spectrum* of $f(t)$. The spectrum for $f(t) = \sin t$ is shown in Figure 7.5, where we have plotted the b_n coefficients versus n. In this case, there is only one nonzero point on the graph.
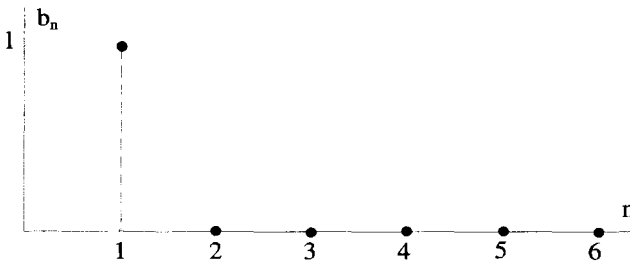
Figure 7.5   The Spectrum for $f(t) = \sin t$

Example 7.2   As a second example, let $f(t)$ be the periodic triangular waveform shown in Figure 7.6. Notice that this function has the same amplitude and period as the previously discussed sine wave. On the interval $-\pi/2 < t < 3\pi/2$, this triangular wave can be expressed as

$$f_T(t) = \begin{cases} 2t/\pi & -\pi/2 \leq t \leq \pi/2 \\ 2 - 2t/\pi & \pi/2 \leq t \leq 3\pi/2 \end{cases}. \tag{7.20}$$

As before, the basic period of this signal is $T_o = 2\pi$, so $\omega_n = n$. Because this is an odd function of t, we can immediately see that all the a_n coefficients are zero. The b_n
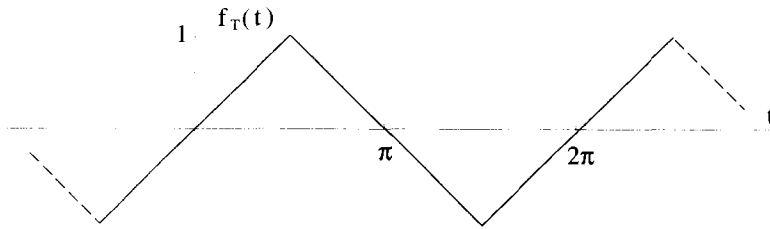
Figure 7.6 The Triangle Wave

coefficients can be calculated from

$$b_n = \frac{1}{\pi} \int_{-\pi/2}^{3\pi/2} dt \ f_T(t) \sin(nt)$$

$$= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} dt \ \frac{2t}{\pi} \sin(nt) + \frac{1}{\pi} \int_{\pi/2}^{3\pi/2} dt \ \left(2 - \frac{2t}{\pi}\right) \sin(nt), \qquad (7.21)$$

with the result that

$$b_n = \frac{8}{\pi^2 n^2} \sin \frac{n\pi}{2}. \qquad (7.22)$$

A plot of the spectrum for the triangle wave is shown in Figure 7.7.
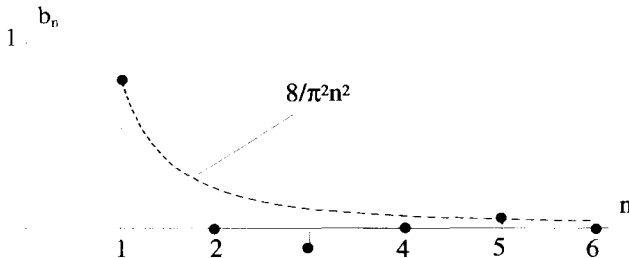
Figure 7.7 The Spectrum for the Triangle Wave

Example 7.3 As a final example, let's calculate the Fourier series for the square wave function $f_S(t)$, shown in Figure 7.8. If this function has unit amplitude and a period of $T_o = 2\pi$, it can be expressed as

$$f_S(t) = \begin{cases} +1 & 0 \le t < \pi \\ -1 & \pi \le t < 2\pi \end{cases}. \qquad (7.23)$$

Again, this is an odd function of t and so all the a_n coefficients are zero. The b_n coefficients can be evaluated from Equation 7.18:
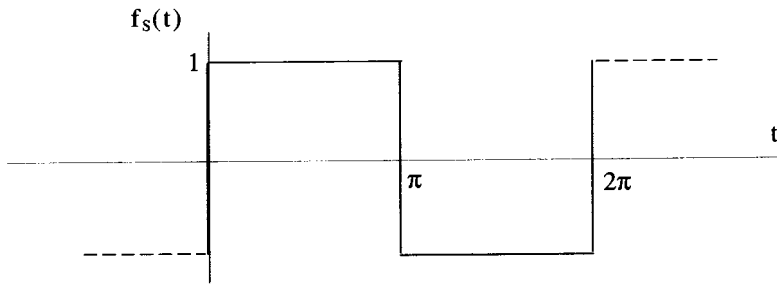
Figure 7.8   The Square Wave

$$b_n = \frac{1}{\pi} \int_0^{2\pi} dt\ f_T(t) \sin(nt)$$

$$= \frac{1}{\pi} \int_0^{\pi} dt \sin(nt) - \frac{1}{\pi} \int_\pi^{2\pi} dt\ \sin(nt), \qquad (7.24)$$

with the result

$$b_n = \begin{cases} \frac{4}{n\pi} & n = 1, 3, 5, 7, \ldots \\ 0 & n = 2, 4, 6, 8, \ldots \end{cases}. \qquad (7.25)$$

The spectrum for the square wave is shown in Figure 7.9.

By comparing Figures 7.5, 7.7, and 7.9, you will notice that as the waveform changes from a sine wave, to a triangular wave, and finally to a square wave, the relative amplitudes of the higher frequency components increase.

The convergence of the Fourier series to the square wave function is demonstrated in Figure 7.10. Three plots are shown. In the first, just the $n = 1$ term of the series is shown. The middle plot shows the result of the sum of the $n = 1$ and $n = 3$ terms. The last plot on the right is the sum of the $n = 1$, $n = 3$, and $n = 5$ terms. Notice that, while the square wave itself is not well defined at its discontinuities, its Fourier
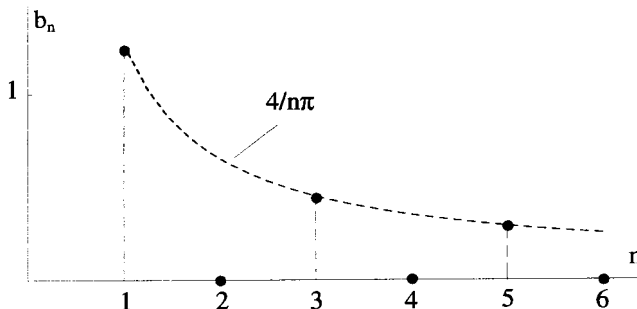
Figure 7.9   The Spectrum for the Square Wave

$$n = 1 \qquad\qquad \sum_{n=1}^{3} \qquad\qquad \sum_{n=1}^{5}$$
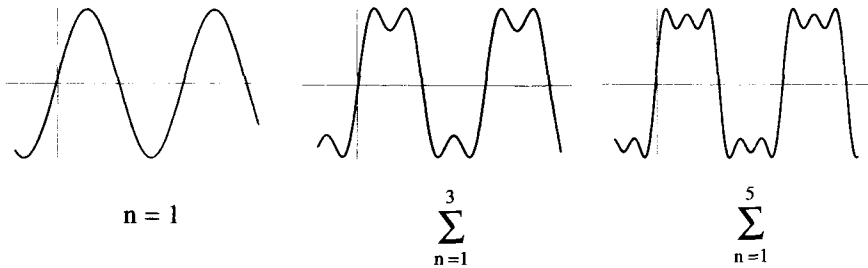
Figure 7.10   The Convergence of the Fourier Series to a Square Wave

series converges to a value equal to the average of the step at these points. There is another interesting convergence effect going on at this discontinuity called the Gibbs phenomenon, which will be discussed later in this chapter.

7.2 THE EXPONENTIAL FORM OF FOURIER SERIES

It turns out that representing the Fourier series with sines and cosines is not very convenient. This is because the sine and cosine terms need to be manipulated separately. There is an alternative representation of the Fourier series, the complex exponential form, which combines the two sets of terms into one series. We can convert the Fourier series of Equation 7.1 into exponential form using the identities

$$\cos\theta = \frac{1}{2}\left(e^{i\theta} + e^{-i\theta}\right) \tag{7.26}$$

$$\sin\theta = \frac{1}{2i}\left(e^{i\theta} - e^{-i\theta}\right). \tag{7.27}$$

7.2.1 The Exponential Series

The exponential form of the Fourier series follows directly from the sine-cosine form

$$s(t) = \frac{a_o}{2} + \sum_{n=1}^{\infty} a_n \cos(\omega_n t) + \sum_{n=1}^{\infty} b_n \sin(\omega_n t). \tag{7.28}$$

Using Equations 7.26 and 7.27, we can write

$$a_m \cos(\omega_m t) + b_m \sin(\omega_m t) = \frac{a_m}{2}\left(e^{i\omega_m t} + e^{-i\omega_m t}\right) + \frac{b_m}{2i}\left(e^{i\omega_m t} - e^{-i\omega_m t}\right). \tag{7.29}$$

The RHS of this expression can be rearranged to give

$$a_m \cos(\omega_m t) + b_m \sin(\omega_m t) = \left(\frac{a_m}{2} + \frac{b_m}{2i}\right)e^{i\omega_m t} + \left(\frac{a_m}{2} - \frac{b_m}{2i}\right)e^{-i\omega_m t}. \tag{7.30}$$

Equation 7.28 can then be placed in the form

$$s(t) = \sum_{n=-\infty}^{\infty} \underline{c}_n e^{i\omega_n t}, \tag{7.31}$$

where

$$c_n = \begin{cases} a_o/2 & n = 0 \\ (a_n - ib_n)/2 & n > 0 \\ (a_{-n} + ib_{-n})/2 & n < 0 \end{cases} . \tag{7.32}$$

Notice that if all the a_n and b_n coefficients are real, Equations 7.32 imply $\underline{c}_n = \underline{c}^*_{-n}$. In that case, the exponential form can also be written

$$s(t) = c_0 + 2 \, \text{Real} \sum_{n=1}^{\infty} \underline{c}_n e^{i\omega_n t}. \tag{7.33}$$

7.2.2 Orthogonality Conditions for Exponential Terms

The $\underline{c}_n$ coefficients can always be obtained by determining the a_n and b_n coefficients using the techniques already described. Better yet, they can be determined directly using a new orthogonality condition:

$$\int_{t_o}^{t_o+T_o} dt \, e^{i\omega_n t} e^{-i\omega_m t} = \delta_{nm} T_o. \tag{7.34}$$

This condition follows directly from the definition of the exponential function and the previous sine and cosine orthogonality conditions. It is true for any t_o, n, and m. The period T_o of $f(t)$ is still related to the fundamental frequency by $T_o = 2\pi/\omega_o$.

If a function $f(t)$ is to be expanded in an exponential Fourier series

$$f(t) = \sum_{n=-\infty}^{\infty} \underline{c}_n e^{i\omega_n t}, \tag{7.35}$$

the $\underline{c}_n$ coefficients can be determined by using the orthogonality condition in Equation 7.34. To do this, multiply both sides of Equation 7.35 by $e^{-i\omega_m t}$ and integrate over the period, T_o. This gives

$$\int_{t_o}^{t_o+T_o} dt \, f(t) e^{-i\omega_m t} = \int_{t_o}^{t_o+T_o} dt \sum_{n=-\infty}^{\infty} \underline{c}_n e^{i\omega_n t} e^{-i\omega_m t}. \tag{7.36}$$

Exchanging the summation and integration on the RHS of this expression and using the orthogonality condition gives:

$$\int_{t_o}^{t_o+T_o} dt \, f(t) e^{-i\omega_m t} = \sum_{n=-\infty}^{\infty} \delta_{mn} \underline{c}_n T_o$$

$$= \underline{c}_m T_o. \tag{7.37}$$

Therefore, we can evaluate the $\underline{c}_n$ coefficient using the expression

$$\underline{c}_n = \frac{1}{T_o} \int_{t_o}^{t_o+T_o} dt\ f(t)e^{-i\omega_n t}, \tag{7.38}$$

which works for all the $\underline{c}_n$'s, including $\underline{c}_o$.

7.2.3 Properties of the Exponential Series

There are three important properties of the exponential form of the Fourier series that can simplify calculations. The first deals with the $\underline{c}_o$ term of the expansion. Because

$$\underline{c}_o = \frac{1}{T_o} \int_0^{T_o} dt\ f(t), \tag{7.39}$$

$\underline{c}_o$ is just the average value of $f(t)$ over the period. This is often written as

$$\underline{c}_o = \langle f(t) \rangle. \tag{7.40}$$

If $f(t)$ is a real function, $\underline{c}_o$ must be a pure real number.

The even or odd nature of $f(t)$ places conditions on the $\underline{c}_n$. If $f(t)$ is a real, even function, then $f(\alpha) = f(-\alpha)$, and

$$\underline{c}_n = \frac{1}{T_o} \int_0^{T_o} dt\ f(t)e^{-i\omega_n t}$$

$$= \frac{1}{T_o} \int_0^{T_o} dt\ f(-t)e^{-i\omega_n t}. \tag{7.41}$$

If a change of variables is introduced, with $\alpha = -t$ and $d\alpha = -dt$, this equation becomes

$$\underline{c}_n = -\frac{1}{T_o} \int_0^{-T_o} d\alpha\ f(\alpha)e^{i\omega_n \alpha}$$

$$= \frac{1}{T_o} \int_{-T_o}^0 d\alpha\ f(\alpha)e^{i\omega_n \alpha}. \tag{7.42}$$

The integrand $f(\alpha)e^{i\omega_n \alpha}$ is periodic in α with a period of T_o. Therefore,

$$\underline{c}_n = \frac{1}{T_o} \int_0^{T_o} d\alpha\ f(\alpha)e^{i\omega_n \alpha}. \tag{7.43}$$

Now take the complex conjugate of both sides of this equation to get

$$\underline{c}_n^* = \frac{1}{T_o} \int_0^{T_o} d\alpha\ f^*(\alpha)e^{-i\omega_n \alpha}. \tag{7.44}$$

If $f(t)$ is a real function of t, then $f^*(t) = f(t)$, and we obtain the result

$$\underline{c}_n = \underline{c}_n^*. \tag{7.45}$$

Therefore, if $f(t)$ is a real, even function of t, then the coefficients of its exponential Fourier series are all pure real. The Fourier series expressed with sines and cosines

actually has an analogous symmetry. A real, even function expressed in that form will not contain any sine terms. Similar arguments show that if $f(t)$ is a real, odd function, then $\underline{c}_n^* = -\underline{c}_n$, and the coefficients are pure imaginary. The analogous statement for the Fourier series expressed in terms of sines and cosines is that any real, odd function will not contain any cosine terms. If $f(t)$ possesses neither odd nor even symmetry, then the coefficients of the exponential series are complex. In this case the sine/cosine series is made up of both sine and cosine functions.

7.2.4 Convenience of the Exponential Form

A major advantage of the exponential form of the Fourier series becomes evident when it is used with differential equations. If sines and cosines are used instead of the exponential notation, you must keep track of two sets of series, which can become horribly intertwined when derivatives are taken. With the exponential notation, all the terms are treated on equal footing, so the bookkeeping is much easier.

Consider a nonhomogeneous, linear differential equation with constant coefficients:

$$\left(A\frac{d^2}{dt^2} + B\frac{d}{dt} + C\right) f(t) = d(t), \tag{7.46}$$

where $d(t)$ is a known function, and the problem is to find $f(t)$. If $d(t)$ is periodic, it can be represented by a Fourier series

$$d(t) = \sum_{n=-\infty}^{\infty} \underline{D}_n e^{i\omega_n t}, \tag{7.47}$$

with

$$\underline{D}_n = \frac{1}{T_o} \int_0^{T_o} dt \, d(t) e^{-i\omega_n t}. \tag{7.48}$$

The coefficients $\underline{D}_n$ and the frequencies ω_n are both known quantities. The unknown function $f(t)$ can be set equal to its Fourier series

$$f(t) = \sum_{n=-\infty}^{\infty} \underline{F}_n e^{i\omega_n t}. \tag{7.49}$$

Here, the $\underline{F}_n$ are unknown, but the ω_n are the same frequencies used in Equation 7.48, which were determined by the periodicity of $d(t)$. This will always be the case for linear differential equations. For nonlinear equations, $f(t)$ can contain frequencies which are not in $d(t)$.

Substituting Equations 7.47 and 7.48 into the differential equation converts it into an algebraic equation, because all the dependencies on the independent variable t cancel out:

$$\left(-\omega_n^2 A + i\omega_n B + C\right) \underline{F}_n = \underline{D}_n. \tag{7.50}$$

Equation 7.50 can be solved for the unknown coefficients $\underline{F}_n$ and used to express the solution in terms of a Fourier series:

$$\underline{F}_n = \frac{-\underline{D}_n}{A\omega_n^2 - i\omega_n B - C} \tag{7.51}$$

$$f(t) = \sum_{n=-\infty}^{\infty} \left(\frac{-\underline{D}_n}{A\omega_n^2 - i\omega_n B - C} \right) e^{i\omega_n t}. \tag{7.52}$$

If sines and cosines had been used to expand the Fourier series for $d(t)$ and $f(t)$, the time dependence could not have been removed as easily. It might be a worthwhile exercise to try this for yourself. The Fourier series approach outlined in Equations 7.49–7.52 will be described in more detail in Chapter 10, when we discuss solution methods for differential equations.

7.3   CONVERGENCE OF FOURIER SERIES

Up to this point we have assumed that a periodic $f(t)$ could always be expressed with an exponential series

$$f(t) = \sum_{n=-\infty}^{\infty} \underline{c}_n e^{i\omega_n t}, \tag{7.53}$$

with $\underline{c}_n$ coefficients given by

$$\underline{c}_n = \frac{1}{2\pi} \int_0^{T_o} dt \, f(t) e^{-i\omega_n t}. \tag{7.54}$$

Actually, the equivalence of $f(t)$ and the series on the RHS of Equation 7.53 was just an assertion that we must justify by demonstrating that the Fourier series converges to the function $f(t)$. In fact, this will not be possible for all types of functions.

This section presents the conditions necessary for the convergence of Fourier series and introduces some new terminology. The proofs of the statements in this section are too involved for this short chapter. We present only a few important results here without proving them. A complete and rigorous reference for this material is the book *Fourier Series* by Georgi Tolstov.

7.3.1   Types of Convergence

In general, to show convergence, one starts with a function, $f(t)$, and calculates the Fourier coefficients $\underline{c}_n$ using Equation 7.54. This step actually places a first, obvious condition on what types of functions can be represented by a Fourier series. It must be possible to perform the integration of Equation 7.54. Next, a version of the Fourier series with a finite number of terms is constructed using these coefficients.

Convergence is proved by showing as the number of terms increases, the series gets arbitrarily close, by some measure, to $f(t)$. There are several different types of convergence. We will discuss three in this section: *Uniform*, *Pointwise*, and *Mean-Squared* convergence.

Pointwise convergence is simple to understand. Define the partial Fourier series $S_N(t)$ as

$$S_N(t) \equiv \sum_{n=-N}^{n=+N} \underline{c}_n e^{i\omega_n t}. \qquad (7.55)$$

This sequence of functions is said to converge pointwise to the function $f(t)$ in the range $a \leq t \leq b$ if

$$\lim_{N \to \infty} [S_N(t) - f(t)] = 0 \qquad (7.56)$$

for any given point t in that range.

The same series is said to converge to $f(t)$ in a mean-squared sense if

$$\lim_{N \to \infty} \int_a^b dt \, |S_N(t) - f(t)|^2 = 0. \qquad (7.57)$$

This is a less strict condition. Notice the sequence $S_N(t) = t^N$ converges to the function $f(t) = 0$ over the range $0 \leq t \leq 1$ in the mean-squared sense, but not in a pointwise sense, because of the behavior of the function at $t = 1$.

Uniform convergence is the most strict convergence of the three. The sequence $S_N(t)$ converges uniformly to the function $f(t)$ on the range $a \leq t \leq b$ if

$$\lim_{N \to \infty} |S_N(t) - f(t)|_{\text{max}} = 0, \qquad (7.58)$$

where max gives the largest value on the range.

7.3.2   Uniform Convergence for Continuous Functions

We state the following theorem without proof. If a periodic function is continuous and piecewise smooth, its Fourier series will always converge uniformly everywhere. A function is piecewise smooth if its derivative exists everywhere, except at isolated points. The triangle wave shown in Figure 7.6 is continuous and piecewise smooth, but the square wave of Figure 7.8 is not.

7.3.3   Mean-Squared Convergence for Discontinuous Functions

If a function is discontinuous, the Fourier series does not converge uniformly, but it may still converge in a mean-squared sense. A good example of a function with a discontinuity is the square wave function we discussed in an earlier example:

$$f(t) = \begin{cases} +1 & 0 \le t < \pi \\ -1 & \pi \le t < 2\pi \end{cases}. \tag{7.59}$$

The sum of the first three terms of the Fourier series was shown progressively in Figure 7.10. Figure 7.11 shows the transition region for the summation of $N = 40$, 60, and 200 terms of the partial Fourier series. The are two complications that arise with the convergence of the Fourier series at points of discontinuity. The first problem is fairly obvious, while the second problem is quite subtle.

The first problem is evident when you consider the point $t = 0$. According to the definition of $f(t)$ in Equation 7.59, $f(0) = 1$. Clearly, looking at Figures 7.10 and 7.11 shows the Fourier series does not converge to this value, even as $N \to \infty$, but rather to the average value of the step. In general, if a periodic function has a discontinuity where the function jumps from f_1 to f_2, its Fourier series will converge to a value of

$$\frac{f_1 + f_2}{2} \tag{7.60}$$

at the discontinuity.

The second complication of the convergence at the discontinuity is less obvious. Again consider the summation of the first three terms of the Fourier series for the square wave, shown in Figure 7.10. As N is increased from 1 to 2 and then to 3, the partial sum $S_N(t)$ looks more and more like a square wave, but there consistently is an oscillation before and after the rising step. You might expect this oscillation to shrink in size as the number of terms in the sum increases, but this does not happen. In Figure 7.11, notice the overshoot recovers faster as N increases, but its amplitude remains constant at a value around 1.17.

This overshoot of the Fourier series at a discontinuity is referred to as the Gibbs phenomenon. In the region surrounding the discontinuity, the series is converging point by point, but not uniformly. Given any point $t = t_o$, except exactly the point of discontinuity, we can increase N until $S_N(t_o) - f(t_o)$ is arbitrarily close to zero. We cannot, however, decrease $|S_N(t) - f(t)|_{max}$ below the value of 1.17 by increasing N.

This leads us to a more general form of the convergence theorem of the previous section. If a function $f(t)$ is piecewise smooth and piecewise continuous, the Fourier
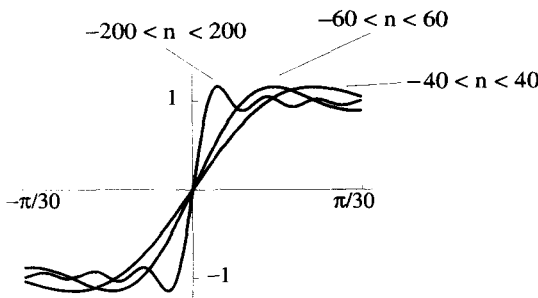
Figure 7.11   The Gibbs Phenomenon

series converges pointwise to the value of $f(t)$ at all points of continuity, and to the mean value of the step of any points of discontinuity. As a whole, the entire series converges to $f(t)$ uniformly if the function is continuous, but only in a mean-squared sense if it has discontinuities.

7.3.4 Completeness

Mathematicians often like to phrase these convergence theorems using the idea of *completeness*. A set of functions is said to form a complete set over a functional class if a linear combination of them can express any of the functions in that class. The complete functions are often called *basis functions*, because they span the space of possible functions, much like the geometric basis vectors span all the possible vectors in space.

Do the exponential terms of the Fourier series form a complete set? In the discussion above, we have said that they form a complete set over the class of piecewise smooth, continuous functions if we require uniform convergence. The theorem in the previous section shows they also form a complete set over the class of piecewise continuous, piecewise smooth functions if we only require mean-squared convergence. Notice we must always declare what kind of convergence is being used before we can say whether a collection of basis functions is complete.

7.4 THE DISCRETE FOURIER SERIES

The coefficients of the Fourier series, either the a_n and b_n of Equation 7.1 or the c_n of Equation 7.31, describe the frequency spectrum of a signal. Often, in physical experiments, there is an electrical signal which we would like to analyze in the frequency domain. Before the advent of the modern, digital era, electrical spectra were obtained using expensive analog spectrum analyzers. In simplified terms, these used an array of narrow-band, analog filters which split incoming signals into different frequency "bins" which were then reported on the output of the device. Modern frequency analyzers obtain their spectra by digitally recording the incoming signal at discrete time intervals, and then using a computer algorithm called the *Fast Fourier Transform* (FFT).

When a computer is used to perform the Fourier series analysis of a signal, it cannot exactly carry out the operations of the previous sections. Because a computer cannot deal with either continuous functions or infinite series, integrals must be replaced by discrete sums, and all summations must be taken over a finite number of terms. The following sections describe a simplified version of the FFT, which adjusts our previous methods in an appropriate manner for computer calculations.

7.4.1 Development of the Discrete Series Equations

The first step in the development of a discrete Fourier series for $f(t)$ is to determine T_o, the period of the signal. Next the computer takes $2N$ samples of $f(t)$, spread
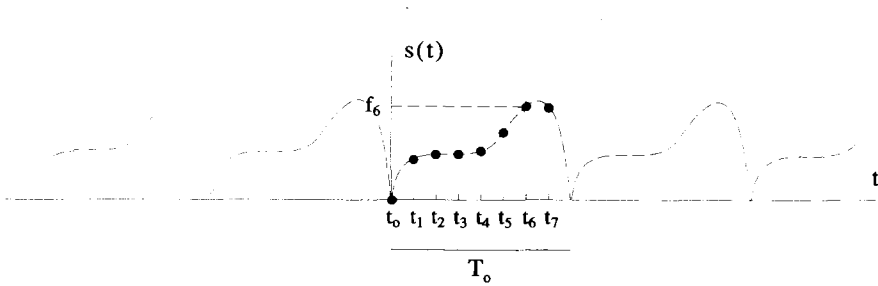
Figure 7.12   The Sampling of $f(t)$ $2N = 8$ Times

evenly over the period, to generate the values

$$f_k = f(t_k) \qquad t_k = \frac{T_o k}{2N} \quad \text{for } k = 0, 1, 2, \dots 2N - 1, \tag{7.61}$$

as shown in Figure 7.12. The first sample f_0 is taken at $t_0 = 0$, and the last sample f_{2N-1} is taken at $t_{2N-1} = T_o(2N - 1)/(2N)$.

The standard exponential Fourier series is summarized by the pair of equations:

$$f(t) = \sum_{n=-\infty}^{\infty} \underline{c}_n e^{i\omega_n t} \tag{7.62}$$

$$\underline{c}_n = \frac{1}{T_o} \int_0^{T_o} dt \, f(t) e^{-i\omega_n t}, \tag{7.63}$$

where $\omega_n = 2\pi n/T_o$. These equations need to be modified for the computer. The new versions of Equations 7.62 should generate the sampled values f_k from a new set of $\underline{c}_n$, via some kind of finite sum:

$$f(t) = \sum_{n=-\infty}^{\infty} \underline{c}_n e^{i\omega_n t} \rightarrow f_k = \sum_{n=0}^{n_{max}} \underline{c}_n e^{i\omega_n t_k}. \tag{7.64}$$

In Equation 7.64, ω_n is defined, as it was before, as $\omega_n = (2\pi n)/T_o$, and n is an integer which we will arbitrarily start at zero and let run to a currently undefined, upper limit n_{max}. The continuous variable t has been replaced by t_k, the set of discretely sampled times.

To evaluate the $\underline{c}_n$ of Equation 7.64, an orthogonality condition similar to Equation 7.34 is needed. We can try to construct one by converting the integral over t into a sum over all the t_k values:

$$\int_0^{T_o} dt \, e^{i\omega_n t} e^{-i\omega_m t} = T_o \delta_{nm} \rightarrow \sum_{k=0}^{2N-1} e^{i\omega_n t_k} e^{-i\omega_m t_k} = ?. \tag{7.65}$$

Notice that the summation over the discrete time index k, unlike the frequency index n, has a well-defined upper limit determined by the number of sample points. The

question now is, does the sum over k in Equation 7.65 lead to an orthogonality condition that allows a determination of the c_n's in Equation 7.64? Answering this question will take some work.

Equation 7.65 can be rewritten as

$$\sum_{k=0}^{2N-1} e^{i\omega_n t_k} e^{-i\omega_m t_k} = \sum_{k=0}^{2N-1} e^{i\frac{\pi}{N}(n-m)k}, \tag{7.66}$$

where we have substituted $\omega_n = 2\pi n/T_o$. Defining

$$r \equiv e^{i\frac{\pi}{N}(n-m)}, \tag{7.67}$$

the sum in the RHS of Equation 7.66 can be recognized as a geometric series, which can be expressed in closed form as

$$\sum_{k=0}^{2N-1} r^k = \frac{1 - r^{2N}}{1 - r}. \tag{7.68}$$

The summation in Equation 7.66 can therefore be written as

$$\sum_{k=0}^{2N-1} e^{i\omega_n t_k} e^{-i\omega_m t_k} = \frac{1 - e^{i2\pi(n-m)}}{1 - e^{i\frac{\pi}{N}(n-m)}}. \tag{7.69}$$

Because n and m are both integers, the numerator of Equation 7.69 is always zero. The LHS of this equation is therefore zero unless the denominator is also zero. This occurs when

$$n - m = 0, \pm 2N, \pm 4N, \ldots. \tag{7.70}$$

From our original definition, the integers n and m range from zero to the unspecified limit n_{max} of the sum in Equation 7.64:

$$0 \leq n \leq n_{max}$$

$$0 \leq m \leq n_{max}. \tag{7.71}$$

If we limit the sum in Equation 7.64 to $2N$ terms, that is $n_{max} = 2N - 1$, we can force

$$|n - m| \leq 2N - 1. \tag{7.72}$$

This important result makes the denominator of Equation 7.69 zero only when $n = m$. This is exactly what we need to construct the orthogonality condition. All that remains is to evaluate Equation 7.69 for $n = m$. We can easily accomplish this by returning to Equation 7.66 and substituting $(n - m) = 0$. Our final orthogonality condition becomes

$$\sum_{k=0}^{2N-1} e^{i\omega_n t_k} e^{-i\omega_m t_k} = \delta_{nm} 2N. \tag{7.73}$$

This orthogonality condition can now be used to invert Equation 7.64 and evaluate the $\underline{c}_n$ for the discrete Fourier series. To do this, we multiply both sides of Equation 7.64 by $e^{-i\omega_m t_k}$ and sum over k:

$$\sum_{k=0}^{2N-1} e^{-i\omega_m t_k} f_k = \sum_{k=0}^{2N-1} e^{-i\omega_m t_k} \sum_{n=0}^{2N-1} \underline{c}_n e^{i\omega_n t_k}$$

$$= \sum_{n=0}^{2N-1} \underline{c}_n \sum_{k=0}^{2N-1} e^{i\omega_n t_k} e^{-i\omega_m t_k}.$$

Using the orthogonality condition gives

$$\sum_{k=0}^{2N-1} e^{-i\omega_m t_k} f_k = \sum_{n=0}^{2N-1} \underline{c}_n \delta_{nm} 2N, \tag{7.74}$$

with the result that

$$\underline{c}_m = \frac{1}{2N} \sum_{k=0}^{2N-1} f_k e^{-i\omega_m t_k}. \tag{7.75}$$

Thus the original equation pair which defines the Fourier series can, for the discrete Fourier series, be replaced by:

$$f(t) = \sum_{n=-\infty}^{\infty} \underline{c}_n e^{i\omega_n t} \rightarrow f_k = \sum_{n=0}^{2N-1} \underline{c}_n e^{i\omega_n t_k}. \tag{7.76}$$

$$\underline{c}_n = \frac{1}{T_o} \int_0^{T_o} dt\, f(t) e^{-i\omega_n t} \rightarrow \underline{c}_n = \frac{1}{2N} \sum_{k=0}^{2N-1} f_k e^{-i\omega_n t_k}. \tag{7.77}$$

7.4.2 Matrix Formulation

Equations 7.76 and 7.77 are easily converted to a matrix format, which in turn, is easily handled by a computer. If we define the elements of the matrix $[M]$ as

$$M_{nk} = e^{i\omega_n t_k} = e^{i\frac{\pi}{N}nk}, \tag{7.78}$$

Equation 7.76 can be written using subscript notation as

$$f_k = \underline{c}_n M_{nk}. \tag{7.79}$$

Likewise, Equation 7.77 becomes

$$\underline{c}_n = \frac{1}{2N} M_{nk}^* f_k. \tag{7.80}$$

The orthogonality condition, Equation 7.73, is simply

$$M_{nk}M_{mk}^* = \delta_{nm}2N. \tag{7.81}$$

Keep in mind, the subscripts of all these expressions run from 0 to $2N - 1$.

Notice that the $[M]$ matrix is the same for all functions, it only depends upon the number of sample points. The computer obtains the coefficients of the discrete Fourier series by inserting the sampled values of $f(t)$ into Equation 7.80 and performing the summations.

Example 7.4   Consider a very simple signal,

$$f(t) = \cos t. \tag{7.82}$$

The standard exponential Fourier series for this signal gives a spectrum with only two nonzero components, $\underline{c}_1 = \underline{c}_{-1} = 1/2$. These components occur at $\omega = \pm 1$, the natural frequency of the signal. A discrete Fourier series can be constructed for this signal by sampling it four times over its 2π period, as shown in Figure 7.13. Table 7.1 shows the sampled values and sampled times.

Any signal sampled four times in a period has $2N = 4$ and an $[M]$ matrix given by

$$M_{nk} = e^{i\frac{\pi}{2}nk}, \tag{7.83}$$
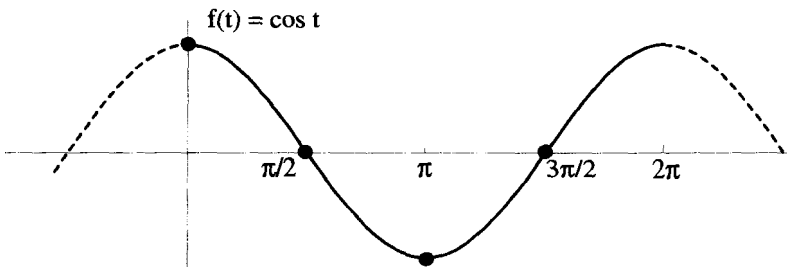
Figure 7.13   The $\cos t$ Function Sampled Four Times

TABLE 7.1.   Sample Times and Values of $\cos t$

k	t_k	f_k
0	0	1
1	$\pi/2$	0
2	π	-1
3	$3\pi/2$	0

which in matrix array notation is

$$[M] = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}. \qquad (7.84)$$

Using Equation 7.80, we can evaluate the c_n,

$$\begin{bmatrix} \underline{c}_0 \\ \underline{c}_1 \\ \underline{c}_2 \\ \underline{c}_3 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \qquad (7.85)$$

with the result that

$$\begin{bmatrix} \underline{c}_0 \\ \underline{c}_1 \\ \underline{c}_2 \\ \underline{c}_3 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}. \qquad (7.86)$$

A spectral plot of these coefficients versus $\omega_n = n$ is shown in Figure 7.14.

Parts of this spectrum make sense, and other parts do not. The fact that $\underline{c}_0$ and $\underline{c}_2$ are zero, while $\underline{c}_1$ is nonzero, agrees with the regular Fourier series spectrum. But the component at $\omega_3 = 3$ makes no sense, because there obviously is no $\omega = 3$ component in $\cos t$. This suspicious extra component is referred to as an *alias*, an effect we will discuss in detail in the next section.
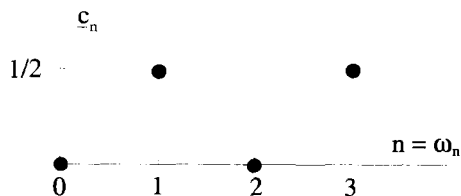
Figure 7.14 Discrete Fourier Series Spectrum for $\cos t$ Sampled Four Times

Now let's see how the coefficients give us back the sampled values of the function. Equation 7.79 written for this example becomes

$$f_k = \sum_{n=0}^{3} \underline{c}_n e^{i\omega_n t_k} = \sum_{n=0}^{3} \underline{c}_n e^{i\frac{\pi}{2}nk}. \tag{7.87}$$

Inserting the values for $\underline{c}_n$ and noting that $t_k = kT_o/(2N) = k\pi/2$,

$$f_k = f(t = t_k) = \frac{1}{2}e^{it_k} + \frac{1}{2}e^{i3t_k}. \tag{7.88}$$

The RHS of Equation 7.88 sums to the proper, real value of $\cos t$ at the sample times $t = 0$, $\pi/2$, π, and $3\pi/2$. At other values of $t \neq t_k$, the RHS of Equation 7.88 does not equal $f(t)$ and is not even pure real! In order for Equation 7.76 to be a good representation of the original $f(t)$, it seems we must sample the signal many times. We will quantify this statement in the following sections and in the next chapter.

7.4.3 Aliasing

In the previous section, the discrete Fourier series technique was applied to the function $f(t) = \cos t$ and it was sampled four times. The resulting spectrum, shown in Figure 7.14, had a legitimate component at $\omega = 1$ and an erroneous one at $\omega = 3$. We called the $\omega = 3$ component an alias. What is the source of this alias, and what parts of a discrete spectrum are to be trusted?

The alias arises from the simple fact that when four sampling points are used, the two functions $\cos t$ and $\cos 3t$ give exactly the same values at the sampling points $(0, \pi/2, \pi, 3\pi/2, 2\pi)$. We can remove the alias at $w = 3$ by sampling more times per period. If $\cos t$ is sampled eight times, instead of four, the new spectrum appears as shown in Figure 7.15. The component at $\omega = 1$ remains unchanged, the component at $\omega = 3$ is now zero, but there is an alias at $\omega = 7$. And, as you probably guessed, the alias occurs because $\cos 7t$ and $\cos t$ have the same values at the eight sampling points $(0, \pi/4, \pi/2, 3\pi/4, \pi, 5\pi/4, 3\pi/2, 7\pi/4, 2\pi)$.

The sampling theorem, which we will prove in the next chapter, says in order to accurately portray a signal by sampling, you must sample at least twice the highest frequency present in the signal. If this is done with a discrete Fourier series, the resulting spectrum can be trusted(i.e., there will be no aliasing) up to half the sampling
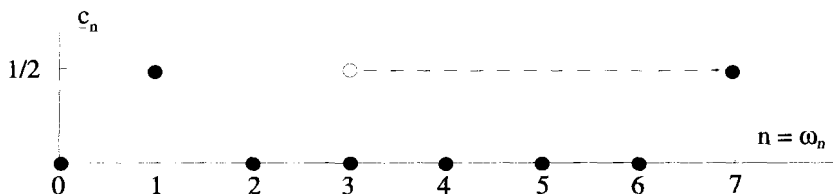
Figure 7.15   The Spectrum of $\cos t$ Sampled Eight Times

frequency. Therefore, the spectrum of Figure 7.14 is valid up to $n = \omega_n = 2$, while the spectrum of Figure 7.15 is valid up to $n = \omega_n = 4$. The signal

$$f(t) = \cos t + \cos 3t \tag{7.89}$$

has a period $T_o = 2\pi$ and must be sampled at a frequency of 6, or six times over that interval, in order to trust the sampled spectrum up to $n = \omega_n = 3$.

7.4.4 Positive and Negative Frequencies

The normal exponential Fourier series contains terms for both positive and negative frequencies. The discrete Fourier series, the way it was developed in the discussion above, created a spectrum that ranged in n from 0 to $(2N - 1)$ or $0 \leq \omega_n \leq 2\pi(2N - 1)/T_o$. To establish this orthogonality condition of Equation 7.73, however, it is only necessary that n take on $2N$ consecutive values. The discrete series analysis can therefore be set up with all the sums over n starting at some arbitrary integer, say n_o, and ranging to $(2N + n_o - 1)$ where $2N$ is still the number of sampled points in the period. Equation 7.76, for example, could have been written as

$$f_k = \sum_{n=n_o}^{2N+n_o-1} \underline{c}_n e^{i2\pi n t_k}. \tag{7.90}$$

The elements of the $[M]$ matrix are generated by the same set of k values, but a shifted set of n values:

$$M_{nk} = e^{i\pi nk/N} \qquad \begin{array}{c} n_o \leq n \leq (2N + n_o - 1) \\ 0 \leq k \leq (2N - 1) \end{array}. \tag{7.91}$$

Except for these modifications, the development of the discrete Fourier series would be unchanged.

This flexibility in selecting the value of n_o, coupled with the sampling theorem, allow discrete Fourier series spectra to be set up more like the regular exponential Fourier spectra. If we set n_o to $-N$, then the sums over n range from $-N$ to $(N - 1)$ and the spectra that result will range in frequency from $-2\pi N/T_o$ to $2\pi(N - 1)/T_o$. If, in addition, the signal is sampled at a rate of at least twice the highest frequency present, there will be no aliasing in the spectrum that results from the discrete Fourier series analysis. Taking such a range for n, the two spectra shown in Figures 7.14 and 7.15 would become as shown in Figure 7.16. In this way a discrete Fourier series spectrum can be generated that more truly represents the real Fourier spectrum of the signal. The crucial step here is to know the highest frequency present. This is sometimes accomplished by passing the signal through a filter that limits the highest frequency to a known value and then sampling at twice this rate or faster.

This way of looking at the discrete Fourier spectrum allows another interpretation of aliasing. The spectral component at $n = 3$ in the spectrum shown in Figure 7.14 can be looked at as an "alias" of the true component at $n = -1$ of the upper spectrum in Figure 7.16. The spectral component at $n = 7$ of Figure 7.15 is an alias of the true component at $n = -1$ of the lower spectrum in Figure 7.16. These spectral
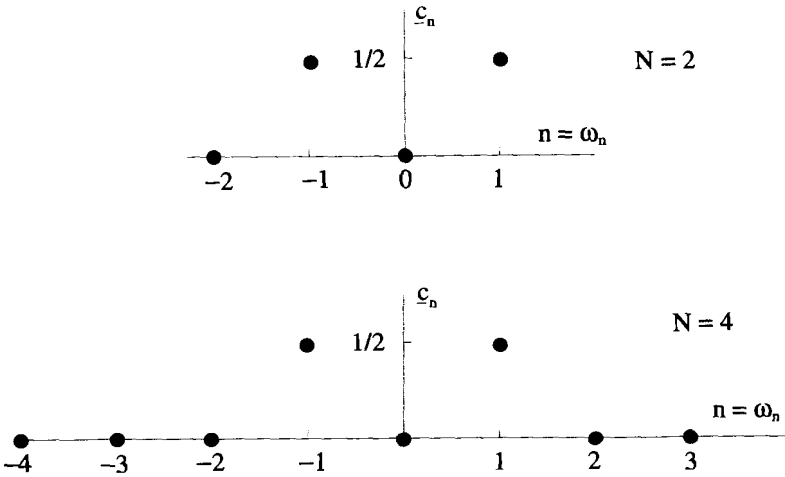
Figure 7.16   The Discrete Fourier Series Spectrum for $-N \le n \le (N - 1)$

components are being moved around by the shifting that occurs in the $[M]$ matrix as n_o is changed, as described by Equation 7.91.

Notice that with the almost symmetric choice for the range of n demonstrated by the spectra in Figure 7.16, the discrete Fourier series representation for the sampled values of $\cos t$ becomes

$$
\begin{aligned}
f_k &= \sum_{n=-N}^{N-1} \underline{c}_n e^{i2\pi n t_k} \\
&= \frac{e^{-it_k}}{2} + \frac{e^{+it_k}}{2} \\
&= \cos t_k.
\end{aligned} \tag{7.92}
$$

The RHS of Equation 7.92 is not only pure real and equal to $\cos t$ at the sampled values, $t = t_k$, but is also pure real and equal to $\cos t$ for all values of t!

EXERCISES FOR CHAPTER 7

1. Show that if $f(t)$ is a periodic function with period T_o, then

$$
f(t) = f(t + 3T_o).
$$

   A Fourier series for $f(t)$ can be constructed that uses $\omega_n = 2\pi n/3T_o$. Show that this is the same Fourier series that is generated using $\omega_n = 2\pi n/T_o$.

2. Determine the coefficients of the Fourier series for the functions $x(t)$, $y(t)$, and $z(t)$ shown below:

x(t)

2

2

y(t)

z(t)
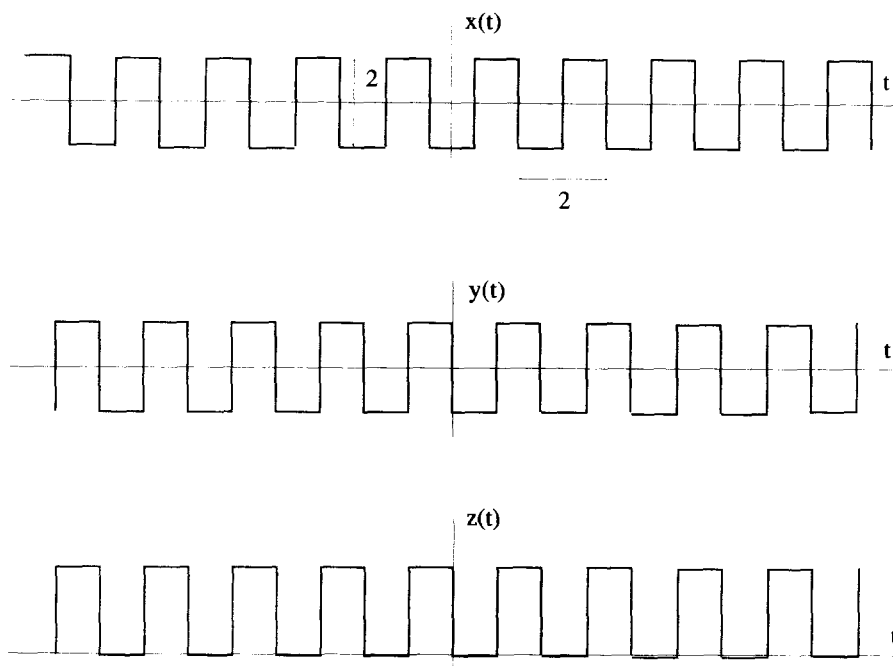
3. Consider the exponential Fourier series for $f(x)$ with the form

$$f(x) = \sum_{n=-\infty}^{\infty} \underline{c}_n e^{i\omega_n x}.$$

Answer the following questions without actually determining the Fourier coefficients:

(a) If $f(x) = \sin^2 x$, what are the values for the ω_n?

(b) If $f(x) = \sin^2 x$, is $\underline{c}_0$ zero or nonzero?

(c) If $f(x) = \sin^2 x$, are the $\underline{c}_n$ pure real, pure imaginary, or complex?

(d) If $f(x) = \sin^2 (x - \pi/2)$, are the $\underline{c}_n$ pure real, pure imaginary, or complex?

(e) If $f(x) = \sin^2 (x - \pi/4)$, are the $\underline{c}_n$ pure real, pure imaginary, or complex?

Now determine all the exponential Fourier series coefficients for $f(x) = \sin^2 x$, *without doing any integrals*.

4. Determine the sine/cosine Fourier series that represents the following periodic function:
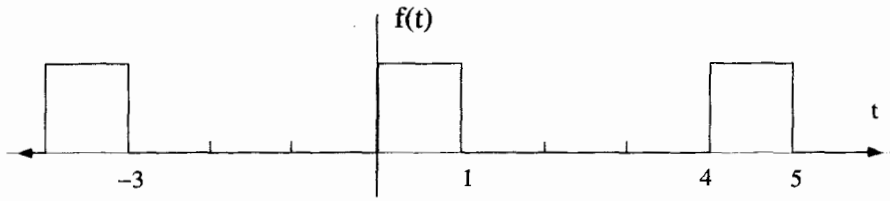
Determine the exponential Fourier series coefficients for this same function.

5. Find the period and coefficients for a Fourier sine/cosine series that represents the following nonperiodic functions over the specified range. Make a plot of the original function and the periodic Fourier series.

   i. $f(t) = t^3$ $0 < t < \pi$
   ii. $f(t) = e^t$ $0 < t < 2\pi$

6. Consider two Fourier series $S_1(x)$ and $S_2(x)$. The first series $S_1(x)$ represents the function $y_1(x) = x^2$ over the range $0 < x < 1$, and $S_2(x)$ is equal to the periodic function $y_2(x)$ where

$$y_2(x) = x \qquad\qquad 0 < x < 1$$
$$y_2(x) = y_2(x + 1) \quad \text{for all } x.$$
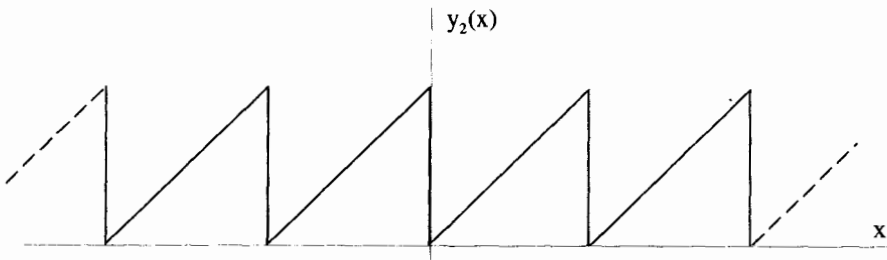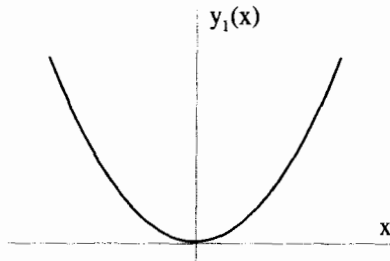
(a) Make a labeled plot of $S_1(x)$ for all x.

(b) Let

$$S_1(x) = \sum_{n=-\infty}^{\infty} a_n e^{ik_n x} \quad \text{and} \quad S_2(x) = \sum_{n=-\infty}^{\infty} b_n e^{ik_n x}.$$

Determine the k_n and, without evaluating the coefficients, find a relationship between the a_n and b_n.

(c) Now evaluate the a_n and b_n coefficients.

7. Consider the periodic function $f(t) = f(t + T_o)$ where

$$f(t) = \begin{cases} 0 & -T_o/2 < t < -\pi \\ \cos t & -\pi < t < +\pi \\ 0 & +\pi < t < +T_o/2 \end{cases}.$$

This function can be expressed as an exponential Fourier series

$$f(t) = \sum_{n=-\infty}^{\infty} \underline{c}_n e^{i\omega_n t}.$$

(a) Make a plot of $f(t)$ vs t.

(b) What are the values for ω_n?

(c) From your answer to part (a) determine $\underline{c}_0$.

(d) Are the $\underline{c}_n$ pure real, pure imaginary, or complex?

(e) Determine the $\underline{c}_n$.

8. Consider the function

$$f(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_o),$$

where n is an integer and $\delta(x)$ is the Dirac δ-function.

(a) Make a labeled plot of $f(t)$ vs t.

(b) Determine the fundamental frequency and the coefficients a_n and b_n of the sine/cosine Fourier series such that

$$\sum_{n=-\infty}^{\infty} \delta(t - nT_o) = \frac{a_o}{2} + \sum_{n=1}^{\infty} a_n \cos(\omega_n t) + \sum_{n=1}^{\infty} b_n \sin(\omega_n t).$$

(c) The δ-function is highly discontinuous, and based on the discussion of convergence in this chapter, we should suspect some odd behavior with the Fourier series representation of this function. What is unusual about the coefficients you determined in part (b)?

9. Evaluate

$$\int_0^{T_o} dt \, \cos(\omega_n t) \, e^{i\omega_m t},$$

where $\omega_n = 2\pi n/T_o$ and n and m are both integers.

10. If a function can be expressed as a Fourier series

$$f(t) = \sum_{n=-\infty}^{\infty} \underline{c}_n e^{i\omega_n t}$$

and its derivative can also be expressed as a Fourier series

$$\frac{df(t)}{dt} = \sum_{n=-\infty}^{\infty} \underline{C}_n e^{i\Omega_n t},$$

how are $\underline{c}_n$ and $\underline{C}_n$ related? How are ω_n and Ω_n related?

11. If $g(t)$ is periodic with period T_o and is represented by a sine/cosine Fourier series with coefficients a_n and b_n, determine the coefficients of the Fourier series that represents the function

$$f(t) = \frac{dg(t)}{dt} + \frac{d^2 g(t)}{dt^2}.$$

Repeat this problem with $g(t)$ represented by an exponential Fourier series.

12. Consider the differential equation

$$\frac{dx(t)}{dt} + x(t) = f(t),$$

where $f(t)$ is a periodic with period T_o and in the interval $0 < t < T_o$ is given by

$$f(t) = \begin{cases} -1 & 0 < t < T_o/2 \\ +1 & T_o/2 < t < T_o \end{cases}. \tag{7.93}$$

(a) Make a plot of $f(t)$.
(b) Determine the exponential Fourier series for $f(t)$.
(c) Assume $x(t)$ can be expressed as an exponential Fourier series and find the coefficients.
(d) What electronic circuit demonstrates the action of this differential equation? What type of electrical signals are represented by $x(t)$ and $f(t)$?

13. Consider the periodic triangular wave shown below:
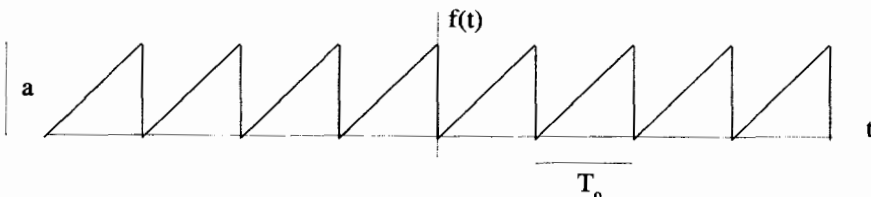
(a) Find the sine/cosine Fourier series for this function.

(b) Find the exponential Fourier series for this function.

(c) Assume that the function $f(t)$ is the driving function for a damped harmonic oscillator, and the response of the oscillator is given by the function $x(t)$ so that

$$m\frac{d^2x(t)}{dt^2} + \lambda\frac{dx(t)}{dt} + kx(t) = f(t).$$

Find the Fourier series for the response $x(t)$. Is it easier to use your answer to part (a) above and find the sine/cosine series for $f(t)$ or your answer to part (b) and find the exponential series for $x(t)$?

(d) What happens to the response $x(t)$ if $T_o = 2\pi\sqrt{m/k}$?

14. Go through the arguments for developing the discrete Fourier series and show that the orthogonality conditions holds as long as the range of n is $n_o \leq n \leq (2N + n_o - 1)$, where $2N$ is the number of samples taken and n_o is any integer.

15. Consider the periodic function $f(t)$ shown below:
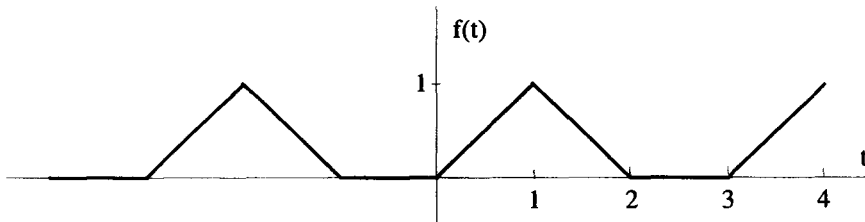
(a) Determine the exponential Fourier series for this function and plot the $\underline{c}_n$ vs. n to describe its spectrum.

(b) Plot the function $df(t)/dt$. Determine its Fourier series and plot its spectrum.

(c) Develop the discrete Fourier series for this function by sampling it six times during its period. Plot the discrete spectrum and discuss the aliasing effects.

16. The function $f(t)$ is the periodic triangular wave shown below.
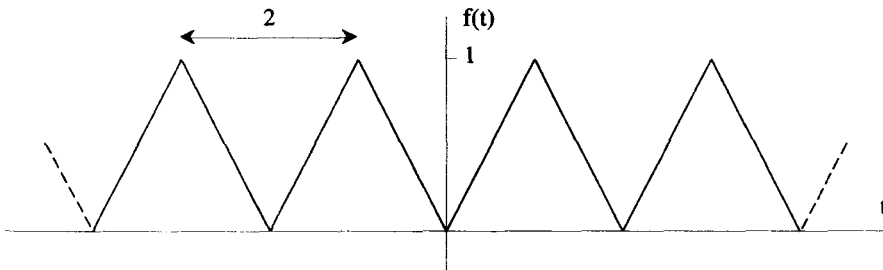
(a) Develop its discrete Fourier series by sampling it four times during a period.

(b) Compare this discrete Fourier series to the one for a cosine function with the same period and sampled at the same rate.

17. Let $f(t) = t^2$ for all t.

(a) Find the exponential Fourier series, $S(t)$ that represents this function over the interval $0 < t < 1$.

(b) Develop the discrete Fourier series for the periodic function $S(t)$ from part (a) above by sampling the function four times during its period. Compare the coefficients found for this discrete Fourier series with the actual ones found in part (a).

(c) Now sample the function $S(t)$ six times in one period and repeat the steps in part (b).

(d) Discuss the aliasing for this problem.

18. Consider a periodic function with $T_o = 4$, given in the range $-1 < t < 3$ by

$$f(t) = \begin{cases} \frac{1}{2}t^2 & -1 < t < +1 \\ -\frac{1}{2}t^2 + 2t - 1 & +1 < t < +3 \end{cases}. \tag{7.94}$$

(a) Make a labeled sketch of $f(t)$.

(b) Find the Fourier series that represents $f(t)$. Note that $d^2 f/dt^2 = g(t)$, where $g(t)$ is a familiar function. Find the Fourier series for $f(t)$ by relating it to the Fourier series for $g(t)$.

(c) Develop the discrete Fourier series for $f(t)$ by sampling it four times during a period. Compare the coefficients of this discrete Fourier series to the coefficients obtained in part (b) above.

(d) Now find the discrete Fourier series by sampling $f(t)$ six times during a period.

19. Let S_N be the partial Fourier series

$$S_N(t) = \sum_{n=-N}^{N} \underline{c}_n e^{i\omega_n t}.$$

The mean-squared error between this partial sum and the function $f(t)$ is defined as

$$\langle \epsilon_N^2 \rangle = \frac{1}{T_o} \int_0^{T_o} dt \, |f(t) - S_N(t)|^2.$$

This error can be looked at as a function of the $2N + 1$ $\underline{c}_n$ coefficients. Assume $S_N(t)$ is pure real and minimize the error with respect to the $\underline{c}_n$, i.e., set

$$\frac{\partial \langle \epsilon_N^2 \rangle}{\partial \underline{c}_n} = 0,$$

to generate $2N + 1$ equations for the $\underline{c}_n$. Show that these equations lead to

$$\underline{c}_n = \frac{1}{T_o} \int_0^{T_o} dt \ f(t) e^{-i\omega_n t},$$

the standard expressions for the Fourier series coefficients.

8

FOURIER TRANSFORMS

The Fourier series analysis of the previous chapter can only be used to obtain spectra for periodic functions. If the independent variable is time, periodic functions must repeat themselves for all time. But signals in the real world have both a beginning and an end, and consequently cannot be analyzed using Fourier series. There is, however, a continuous version of the Fourier series, called the Fourier transform, which can handle certain types of nonperiodic functions. In this chapter, we demonstrate how the Fourier transform is the limiting case of the Fourier series, as the period of repetition grows infinitely large.

8.1 FOURIER SERIES AS $T_o \rightarrow \infty$

The Fourier series for a periodic function, such as the one in Figure 8.1, can be summarized by the equation pair

$$f(t) = \sum_{n=-\infty}^{+\infty} \underline{c}_n e^{i\omega_n t} \tag{8.1}$$

$$\underline{c}_n = \frac{1}{T_o} \int_{-T_o/2}^{T_o/2} dt \, f(t) e^{-i\omega_n t}, \tag{8.2}$$

where $\omega_n = 2\pi n/T_o$. To simplify the discussion that follows, the limits of integration have been shifted to be symmetric about $t = 0$. The $\underline{c}_n$ coefficients are the complex amplitudes of the discrete frequency components which make up $f(t)$.

As the period T_o grows larger, as shown in Figure 8.2, the Fourier series still generates a spectrum, but the discrete frequencies $\omega_n = 2\pi n T_o$ become packed
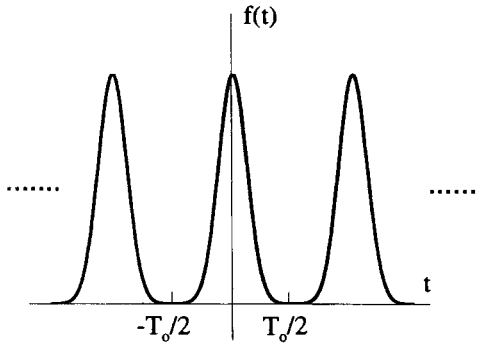
Figure 8.1   Standard Periodic Function

closer and closer together. If T_o is infinite, as shown in Figure 8.3, the function $f(t)$ is no longer periodic. Equation 8.2 can no longer be used to generate a spectrum because of the $1/T_o$ factor, and the Fourier series analysis breaks down.

Functions like the one in Figure 8.3 can be analyzed with the Fourier transform, which can be derived by carefully considering the Fourier series equations in the
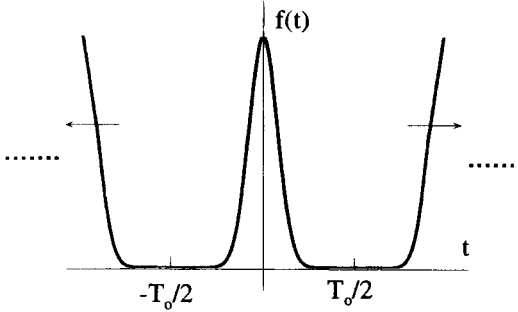
Figure 8.2   Periodic Signal with Increasing T_o
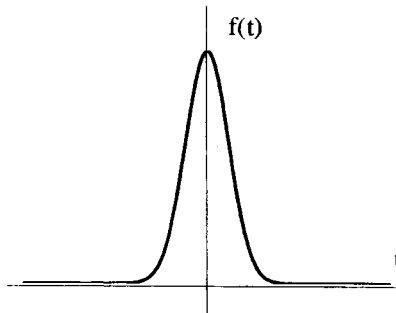
Figure 8.3   Typical Nonperiodic Signal

limit as $T_o \to \infty$. In this limit, the discrete ω_n frequencies merge into a continuous frequency parameter

$$\omega_n = \frac{2\pi n}{T_o} \to \omega \tag{8.3}$$

and we can treat $\Delta \omega$, the step size between the discrete frequencies, as a differential variable $d\omega$

$$\Delta \omega = \frac{2\pi}{T_o} \to d\omega. \tag{8.4}$$

To find the equations equivalent to the Fourier series pair given by Equations 8.1 and 8.2 in the limit of $T_o \to \infty$, insert Equation 8.2 into Equation 8.1 to obtain

$$f(t) = \sum_{n=-\infty}^{\infty} \left[\frac{1}{T_o} \int_{-T_o/2}^{T_o/2} d\tau \, f(\tau) e^{-i\omega_n \tau} \right] e^{i\omega_n t}. \tag{8.5}$$

Notice that in this equation a dummy variable of integration τ has been introduced to avoid confusion with the independent variable t. Rearranging Equation 8.5 and substituting $1/T_o = \Delta \omega / 2\pi$ gives

$$f(t) = \sum_{n=-\infty}^{\infty} \frac{\Delta \omega}{2\pi} e^{i\omega_n t} \left[\int_{-T_o/2}^{T_o/2} d\tau \, f(\tau) e^{-i\omega_n \tau} \right]. \tag{8.6}$$

As T_o goes to infinity, $\omega_n \to \omega$ and $\Delta \omega \to d\omega$, so the sum over n becomes an integral over ω. Thus in this limit, Equation 8.6 becomes

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \, e^{i\omega t} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\tau \, f(\tau) e^{-i\omega \tau} \right], \tag{8.7}$$

where we have arbitrarily split up the factor of $1/2\pi$ in anticipation of things to come.

The bracketed term on the RHS of Equation 8.7 is called the Fourier transform of $f(t)$ and is a function of the continuous variable ω. It is generally a complex function, so we represent it with the symbol $\underline{F}(\omega)$. With this definition for $\underline{F}(\omega)$, Equation 8.7 can be broken into two equations

$$\underline{F}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \, e^{-i\omega t} f(t) \tag{8.8}$$

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \, e^{i\omega t} \underline{F}(\omega), \tag{8.9}$$

which are often referred to as the Fourier transform pair. They are the continuous extensions of Equations 8.1 and 8.2 for nonperiodic functions. These equations are sometimes written using the operator symbol $\mathcal{F}$:

$$\underline{F}(\omega) = \mathcal{F}\{f(t)\} \tag{8.10}$$

$$f(t) = \mathcal{F}^{-1}\{\underline{F}(\omega)\}, \tag{8.11}$$

where

$$\mathcal{F} \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \, e^{-i\omega t} \tag{8.12}$$

and

$$\mathcal{F}^{-1} \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \, e^{i\omega t} \tag{8.13}$$

are both operators.

8.2   ORTHOGONALITY

The orthogonality condition associated with the exponential Fourier series, which allowed the coefficients to be evaluated, took the form

$$\int_0^{T_o} dt \, e^{i(\omega_n - \omega_m)t} = T_o \delta_{nm}, \tag{8.14}$$

where $\omega_n = 2\pi n/T_o$ and n was an integer. A similar orthogonality expression exists for the Fourier transform. As long as the Fourier transform $\underline{F}(w)$ exists, Equation 8.7 can be rearranged as

$$f(t) = \int_{-\infty}^{\infty} d\tau \, f(\tau) \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \, e^{i(t-\tau)\omega} \right]. \tag{8.15}$$

Because the definition of the Dirac delta-function is

$$f(t) = \int_{-\infty}^{\infty} d\tau f(\tau)\delta(t - \tau), \tag{8.16}$$

the bracketed term on the RHS of Equation 8.15 must act like a shifted δ-function:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \, e^{i(t-\tau)\omega} = \delta(t - \tau). \tag{8.17}$$

Interchanging the variables gives the inverted form:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} dt \, e^{i(\omega - \omega')t} = \delta(\omega - \omega'). \tag{8.18}$$

These two equations are the orthogonality conditions we seek. A function $e^{i\omega t}$ is orthogonal to all other functions $e^{-i\omega' t}$ when integrated over all t, as long as $\omega \neq \omega'$.

The relations in Equations 8.17 and 8.18 cannot be viewed as rigorously correct, because they involve a δ-function which is not enclosed in an integral. The relations should be viewed as a shorthand notation and are mathematically valid only if they appear inside integral operations.

The orthogonality relations given in Equations 8.17 and 8.18 can be used to generate one of the equations of the Fourier transform pair (Equations 8.8 and 8.9) from the other. Operating on Equation 8.8 with $\int d\omega\, e^{i\omega\tau}$ and then applying Equation 8.17 gives

$$
\begin{aligned}
\int_{-\infty}^{\infty} d\omega\, e^{i\omega\tau}\underline{F}(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt\, f(t) \int_{-\infty}^{\infty} d\omega\, e^{i(\tau-t)\omega} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt\, f(t)2\pi\delta(\tau - t) \\
&= \sqrt{2\pi}f(\tau),
\end{aligned}
\tag{8.19}
$$

which is simply Equation 8.9. The same process works in reverse by operating on Equation 8.9 with $\int dt\, e^{-i\omega t}$ and using Equation 8.18.

8.3   EXISTENCE OF THE FOURIER TRANSFORM

In order for the Fourier transform to exist, we must be able to perform two processes. First, the transform integral itself must exist. That is, given an $f(t)$, we must be able to calculate $\underline{F}(\omega)$ from Equation 8.8. Second, when we apply the inverse transform, we must retrieve the original function. In other words, after calculating the $\underline{F}(\omega)$, we must be able to apply Equation 8.9 and get back $f(t)$.

In order for the Fourier transform of a function, $f(t)$, to exist, it must be possible to perform the integration,

$$
\int_{-\infty}^{\infty} dt\, e^{-i\omega t} f(t).
\tag{8.20}
$$

One simple condition derives from the fact that

$$
\left|e^{-i\omega t}\right| = 1,
\tag{8.21}
$$

so we can write

$$
\left|\int_{-\infty}^{\infty} dt\, e^{-i\omega t} f(t)\right| \le \int_{-\infty}^{\infty} dt\, |f(t)|.
\tag{8.22}
$$

Thus a sufficient condition for the existence of $\underline{F}(\omega)$ is that the integral on the RHS of Equation 8.22 be finite. That is,

$$
\int_{-\infty}^{\infty} dt\, |f(t)| < \infty
\tag{8.23}
$$

Functions which obey Equation 8.23 are called *absolutely integrable*.

The second issue, whether the inverse transformation returns the appropriate original function, is more subtle. If $\underline{F}(\omega)$ exists, it can be substituted into Equation 8.9 to give Equation 8.15. Then by the orthogonality condition of Equation 8.17, it would appear that if $\underline{F}(\omega)$ exists, the inversion must return the original $f(t)$. This indeed will be the case for all well-behaved, continuous functions. But the sifting action of $\delta(t - \tau)$ is only really well defined for functions that are continuous at $t = \tau$. To determine what happens when the function is discontinuous at $t = \tau$ takes a little more thought.

One way to look at this situation is to substitute a sequence function $\delta_n(t - \tau)$ for the bracketed term in Equation 8.15 and then take the limit as $n \to \infty$. The Fourier inversion equation then becomes

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} d\tau \, f(\tau) \delta_n(t - \tau). \tag{8.24}$$

If a Gaussian sequence function is used, the functions in the integrand can be graphed as shown in Figure 8.4. From this figure, it is clear that in the infinite limit we can write

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} d\tau \, f(\tau) \delta_n(t - \tau) = \lim_{\epsilon \to 0} \frac{f(t - \epsilon) + f(t + \epsilon)}{2}. \tag{8.25}$$

If $f(t)$ is a continuous function, then the RHS of Equation 8.25 becomes $f(t)$ and the Fourier inversion returns the original function. If the function $f(t)$ is discontinuous at some point, say $t = t_o$, the inversion returns the average value of $f(t)$:

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} d\tau \, f(\tau) \delta_n(t - \tau) = \frac{f_l + f_r}{2}, \tag{8.26}$$

where f_l and f_r are the values of $f(t)$ just to the left and right of the discontinuity, as shown in Figure 8.5.

Depending upon how $f(t)$ is defined at the discontinuity, this may or may not be the correct value for $f(t = t_o)$. For example, if $f(t)$ were defined as

$$f(t) = \begin{cases} f_l & t \leq t_o \\ f_r & t > t_o \end{cases}, \tag{8.27}$$
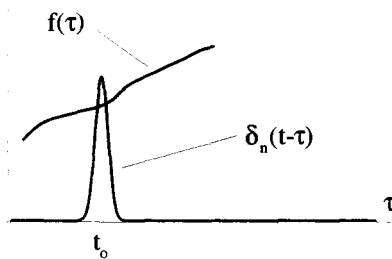
Figure 8.4   Integrand Functions for the Fourier Inversion
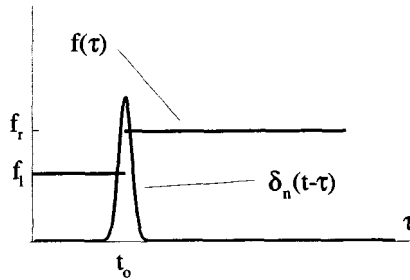
Figure 8.5   Fourier Inversion at a Discontinuity

so that $f(t_o) = f_l$, the Fourier inversion returns the incorrect result given in Equation 8.26. This same effect was seen in the discussion of the convergence of Fourier series in the previous chapter.

8.4   THE FOURIER TRANSFORM CIRCUIT

The expression for $f(t)$ in terms of its Fourier transform $\underline{F}(\omega)$,

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \, e^{i\omega t} \underline{F}(\omega), \tag{8.28}$$

implies that the function $f(t)$ can be made up of a continuous sum of oscillators. This can be seen by considering the integral in Equation 8.28 as the limit of a Riemann sum. Remember that a Riemann sum views an integral, such as the integral of $g(x)$ over all of x, to be the limit of a sum of rectangular areas,

$$\int_{-\infty}^{\infty} dx \, g(x) = \lim_{\Delta x \to 0} \sum_{n=-\infty}^{\infty} g(n\Delta x) \, \Delta x, \tag{8.29}$$

where Δx is the width of each rectangle. Applying this expression to Equation 8.28 gives

$$f(t) = \lim_{\Delta\omega \to 0} \sum_{n=-\infty}^{\infty} \frac{F(n\Delta\omega)}{\sqrt{2\pi}} e^{i(n\Delta\omega)t} \Delta\omega. \tag{8.30}$$

Equation 8.30 represents a sum of terms, each oscillating as $e^{i(n\Delta\omega)t}$, with an amplitude given by $\Delta\omega \underline{F}(n\Delta\omega)/\sqrt{2\pi}$. Similar to what we did in the previous chapter with Fourier series, we can view this sum as the electric circuit shown in Figure 8.6. The value of $f(t)$ is the sum of the voltages produced by all the oscillators. As $\Delta\omega$ goes to zero, both the spacing between the frequencies of the oscillators and the amplitude of each oscillator gets infinitesimally small. The sum, however, combines to form the finite value of $f(t)$.
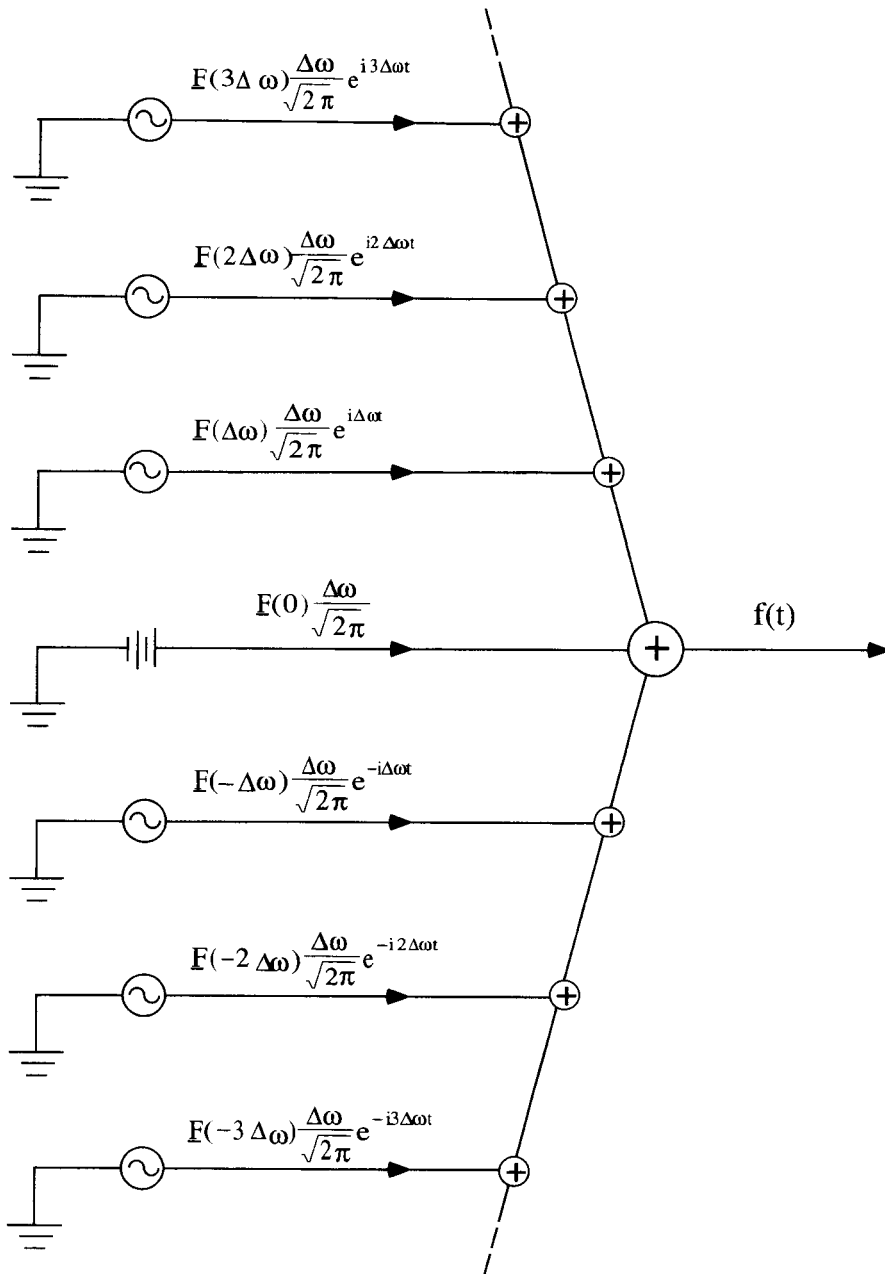
Figure 8.6   The Fourier Transform Circuit

8.5 PROPERTIES OF THE FOURIER TRANSFORM

In this section, a number of mathematical properties of the Fourier transform are developed. Each derivation involves either one or both of the functions $f(t)$ and $g(t)$, which we will assume possess the valid Fourier transforms:

$$\mathcal{F}[f(t)] = \underline{F}(\omega) \tag{8.31}$$

$$\mathcal{F}[g(t)] = \underline{G}(\omega). \tag{8.32}$$

8.5.1 Delay

The Fourier transform of $f(t - t_o)$, which is $f(t)$ delayed by t_o, can be expressed in terms of $\underline{F}(\omega)$. The Fourier transform of the delayed function is

$$\mathcal{F}[f(t - t_o)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt\, e^{-i\omega t} f(t - t_o). \tag{8.33}$$

Make the substitution $\alpha = t - t_o$ to obtain

$$\mathcal{F}[f(t - t_o)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\alpha\, e^{-i\omega(\alpha + t_o)} f(\alpha)$$

$$= e^{-i\omega t_o} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\alpha\, e^{-i\omega\alpha} f(\alpha)$$

$$= e^{-i\omega t_o} \mathcal{F}[f(t)]. \tag{8.34}$$

So when a function is delayed by an amount t_o, its Fourier transform is multiplied by a factor of $e^{-i\omega t_o}$.

8.5.2 Time Inversion of Real Functions

The Fourier transform of $f(-t)$, if $f(t)$ is a real function, is the complex conjugate of $\underline{F}(\omega)$. To see this, start by writing the Fourier transform of the inverted function as

$$\mathcal{F}[f(-t)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt\, e^{-i\omega t} f(-t). \tag{8.35}$$

Substitution of $\alpha = -t$ gives

$$\mathcal{F}[f(-t)] = -\frac{1}{\sqrt{2\pi}} \int_{\infty}^{-\infty} d\alpha\, e^{i\omega\alpha} f(\alpha)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\alpha\, e^{i\omega\alpha} f(\alpha), \tag{8.36}$$

with the result that, if $f(t)$ is pure real,

$$\mathcal{F}[f(-t)] = \underline{F}^{*}(\omega). \tag{8.37}$$

8.5.3 Even and Odd Functions

If $f(t)$ is a real, even function then

$$f(-t) = f(t), \tag{8.38}$$

and by Equations 8.31 and 8.37

$$\underline{F}^*(\omega) = \underline{F}(\omega). \tag{8.39}$$

Thus, for even functions, $\underline{F}(\omega)$ must be a pure real function of ω.

If $f(t)$ is a real, odd function then

$$f(-t) = -f(t), \tag{8.40}$$

and by Equations 8.31 and 8.37

$$\underline{F}^*(\omega) = -\underline{F}(\omega). \tag{8.41}$$

Therefore, for odd functions, $\underline{F}(\omega)$ must be a pure imaginary function of ω.

8.5.4 Spectra for Pure Real $f(t)$

Functions which are pure real have transforms which obey an important symmetry. This can be derived easily by substituting $-\omega$ into Equation 8.8 to give

$$\underline{F}(-\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \, e^{i\omega t} f(t). \tag{8.42}$$

Consequently, if $f(t)$ is pure real then

$$\underline{F}(-\omega) = \underline{F}^*(\omega). \tag{8.43}$$

This means that the negative frequency part of the spectrum can always be obtained from the positive frequency spectrum. The negative frequency part of the spectrum contains no new information.

8.5.5 The Transform of $df(t)/dt$

The Fourier transform of the derivative of $f(t)$ can be expressed in terms of the transform of $f(t)$. To see this, start with the transform of the derivative:

$$\mathcal{F}\left[\frac{df}{dt}\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \, e^{-i\omega t} \frac{df}{dt}. \tag{8.44}$$

The RHS of Equation 8.44 can be integrated by parts. Let $u = e^{-i\omega t}$ and $dv = (df/dt)\,dt$. Then

$$\mathcal{F}\left[\frac{df}{dt}\right] = \frac{1}{\sqrt{2\pi}} \left\{ e^{-i\omega t} f(t) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} dt \, (-i\omega)e^{-i\omega t} f(t) \right\}. \tag{8.45}$$

If $f(t) \to 0$ as $t \to \pm\infty$, which must occur for any function with a valid Fourier transform, then

$$\mathcal{F}\left[\frac{df}{dt}\right] = \frac{i\omega}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \, e^{-i\omega t} f(t)$$

$$= i\omega \underline{F}(\omega). \tag{8.46}$$

8.5.6 The Transformations of the Product and Convolution of Two Functions

The Fourier transform of the product of two functions can be expressed in terms of the Fourier transforms of the individual functions. The Fourier transform of the product of $f(t)$ and $g(t)$ is

$$\mathcal{F}\left[f(t)g(t)\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \, e^{-i\omega t} f(t)g(t). \tag{8.47}$$

To express this transform in terms of $\underline{F}(\omega)$ and $\underline{G}(\omega)$, insert the inverse transforms for $f(t)$ and $g(t)$ into Equation 8.47, making sure to use different variables of integration to keep all the integrals independent:

$$\mathcal{F}\{f(t)g(t)\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \, e^{-i\omega t} \times \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega' \, e^{i\omega' t} \underline{G}(\omega')$$

$$\times \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega'' \, e^{i\omega'' t} \underline{F}(\omega''). \tag{8.48}$$

Now collect all the $1/\sqrt{2\pi}$ factors, and combine the integrals over t, ω' and ω'' into a three-dimensional integral. The RHS of Equation 8.48 becomes

$$\left(\frac{1}{\sqrt{2\pi}}\right)^3 \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\omega' \int_{-\infty}^{\infty} d\omega'' \, [\underline{G}(\omega')\underline{F}(\omega'')e^{i(\omega'+\omega''-\omega)t}]. \tag{8.49}$$

The order of integration is arbitrary. If we choose to do the t integral first, we are confronted with the integral

$$\int_{-\infty}^{\infty} dt \, e^{i(\omega'+\omega''-\omega)t} = 2\pi\delta(\omega' + \omega'' - \omega), \tag{8.50}$$

which is just the orthogonality condition for the Fourier transform. Using this in Equation 8.49 gives

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega' \int_{-\infty}^{\infty} d\omega'' \, [\underline{G}(\omega')\underline{F}(\omega'')\delta(\omega' + \omega'' - \omega)]. \tag{8.51}$$

Performing the integration over $d\omega''$ produces the final result

$$\mathcal{F}\{f(t)g(t)\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega' \, \underline{G}(\omega')\underline{F}(\omega - \omega'). \tag{8.52}$$

The operation on the RHS of Equation 8.52 is a common one and is called the *convolution* of $\underline{G}(\omega)$ and $\underline{F}(\omega)$. Convolution plays an important role in the application of Green's functions for solutions of differential equations, which is discussed in a later chapter. The convolution of $\underline{G}(\omega)$ and $\underline{F}(\omega)$ is often written using the shorthand notation

$$\underline{G}(\omega) \otimes \underline{F}(\omega) \equiv \int_{-\infty}^{\infty} d\omega' \, \underline{G}(\omega')\underline{F}(\omega - \omega'). \tag{8.53}$$

From the form of Equation 8.53, it is clear that

$$\underline{G}(\omega) \otimes \underline{F}(\omega) = \underline{F}(\omega) \otimes \underline{G}(\omega). \tag{8.54}$$

Due to the symmetry of the Fourier transform pair, it is clear that the Fourier inversion of the product of two transforms will also involve convolution:

$$\mathcal{F}^{-1}\{\underline{F}(\omega)\underline{G}(\omega)\} = \frac{1}{\sqrt{2\pi}} f(t) \otimes g(t) \tag{8.55}$$

or

$$\mathcal{F}\{f(t) \otimes g(t)\} = \sqrt{2\pi} \, \underline{F}(\omega)\underline{G}(\omega). \tag{8.56}$$

Convolution can be somewhat difficult to visualize. The process is facilitated by a graphical interpretation demonstrated in the following example.

Example 8.1 Consider the convolution of two functions:

$$h(t) \equiv f(t) \otimes g(t) = \int_{-\infty}^{\infty} d\tau \, f(\tau)g(t - \tau) \tag{8.57}$$

Let $f(t)$ be a square pulse and $g(t)$ a triangular pulse, as shown in Figure 8.7. The value of $h(t)$ at a particular value of t is equal to the area of the integrand of Equation 8.57.
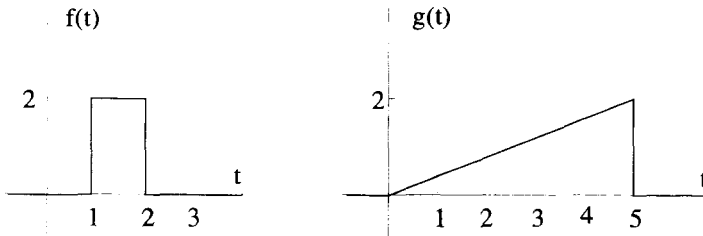
Figure 8.7 Example Functions for Convolution
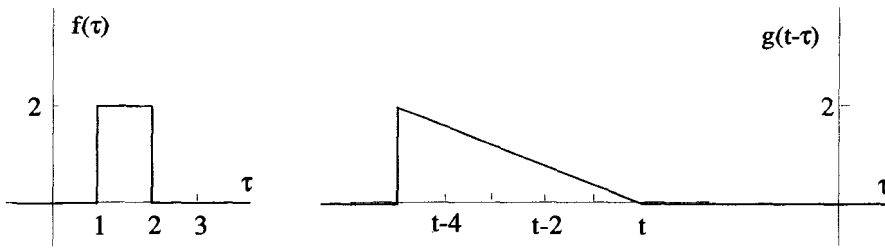
Figure 8.8   Convolution Functions Plotted vs. τ

This area can be represented graphically as the area under the product of $f(\tau)$ and $g(t - \tau)$ plotted vs. τ. The functions $f(\tau)$ and $g(t - \tau)$ plotted vs. τ are shown in Figure 8.8. The function $g(t - \tau)$ plotted vs. τ is the function $g(\tau)$ inverted and shifted by an amount t. In Figure 8.8, $g(t - \tau)$ has been plotted for a negative value of t. The integrand of the convolution integral is a product of these two functions, and the value of the convolution is the area under this product.

To determine $h(t)$, we must consider this process for all possible values of t. For this example, start with $t \ll 0$. A combined plot of $f(\tau)$ and $g(t - \tau)$ for this condition is shown in Figure 8.9, which clearly indicates there is no overlap of the two functions. Their product is therefore zero, and $h(t)$ is zero for this value of t. In fact, from this picture it is clear that $h(t) = 0$ for all $t < 1$ and all $t > 7$. The convolution of these two functions can be nonzero only in the range $1 < t < 7$, where a little more analysis is needed. Within this range there are three subregions of t to consider: $1 < t < 2$; $2 < t < 6$; and $6 < t < 7$. The nature of the overlap between the two functions is different in each case.

For $1 < t < 2$, the functions of Figure 8.9 only overlap in the range $1 < \tau < t$, as shown in Figure 8.10. Therefore for $1 < t < 2$,

$$h(t) = f(t) \otimes g(t) = \int_1^t d\tau\, f(\tau)g(t - \tau). \qquad (8.58)$$

Over the integration range $f(\tau) = 2$. The function $g(t - \tau)$ is given by

$$g(t - \tau) = \begin{cases} (2/5)(t - \tau) & 0 < (t - \tau) < 5 \\ 0 & \text{otherwise} \end{cases}. \qquad (8.59)$$
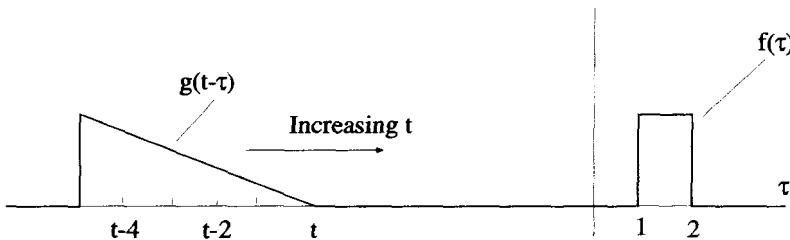
Figure 8.9   Convolution Functions for $t \ll 0$
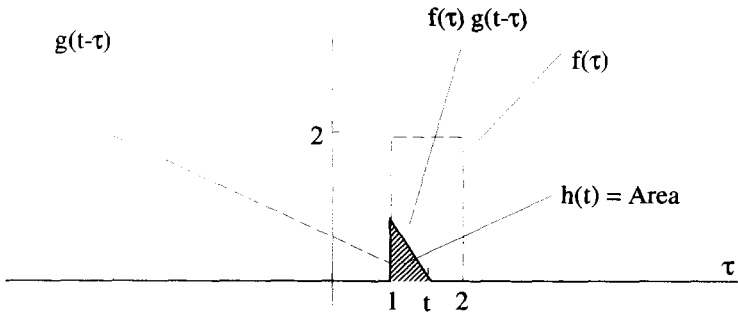
Figure 8.10   Convolution Picture for $1 < t < 2$

The range $0 < (t - \tau) < 5$ is equivalent to $(t - 5) < \tau < t$. Therefore, over the range of the convolution integral in Equation 8.58, $g(t - \tau) = (2/5)(t - \tau)$, and

$$h(t) = f(t) \otimes g(t) = \int_1^t d\tau\,(4/5)(t - \tau) \qquad 1 < t < 2. \qquad (8.60)$$

The result is

$$h(t) = f(t) \otimes g(t) = \frac{2}{5}t^2 - \frac{4}{5}t + \frac{2}{5} \qquad 1 < t < 2. \qquad (8.61)$$

Up to this point, $h(t)$ has been obtained for the ranges $-\infty < t < 2$ and $7 < t$, as shown in Figure 8.11.

For $2 < t < 6$, the triangle function is shifted further to the right, so that the square pulse is totally inside it, as shown in Figure 8.12. The convolution integral in this region becomes

$$h(t) = \int_1^2 d\tau\,f(\tau)g(t - \tau) \qquad 2 < t < 6. \qquad (8.62)$$
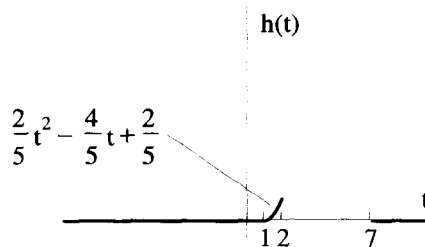
Figure 8.11   Convolution Result for $-\infty < t < 2$ and $7 < t$
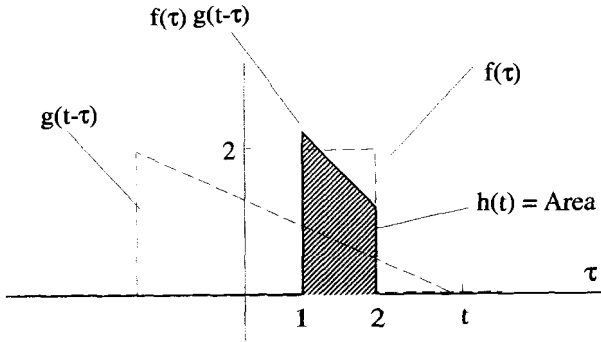
Figure 8.12 Convolution Picture for $2 < t < 6$

As before, $f(\tau) = 2$ and $g(t - \tau) = (2/5)(t - \tau)$, so Equation 8.62 integrates to

$$h(t) = \frac{4}{5}t - \frac{6}{5} \qquad 2 < t < 6. \tag{8.63}$$

The convolution function $h(t)$ has now been determined for the ranges $-\infty < t < 6$ and $7 < t$ as shown in Figure 8.13.

For the last region, $6 < t < 7$, the triangular pulse has moved to the right so only its trailing edge is inside the square pulse, as shown in Figure 8.14. The convolution integral for this range becomes

$$h(t) = \int_{t-5}^{2} d\tau \, f(\tau)g(t - \tau) \qquad 6 < t < 7. \tag{8.64}$$

Again, $f(\tau) = 2$ and $g(t - \tau) = (2/5)(t - \tau)$, so that the integration evaluates to

$$h(t) = -\frac{2}{5}t^2 + \frac{8}{5}t + \frac{42}{5} \qquad 6 < t < 7. \tag{8.65}$$
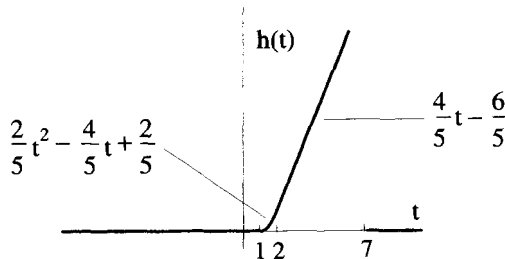
Figure 8.13 Convolution Results for $-\infty < t < 6$ and $7 < t$
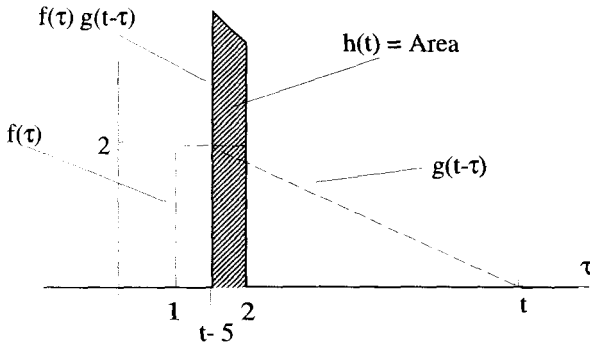
Figure 8.14 Convolution Picture for $6 < t < 7$
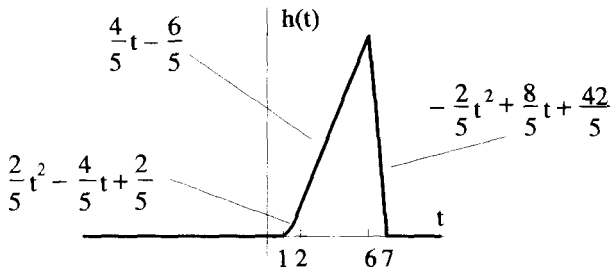
Figure 8.15 Convolution Results for $-\infty < t < \infty$

The final result for the convolution of $f(t)$ with $g(t)$ is shown in Figure 8.15. As you can see from this example, a good deal of care must be taken to perform a convolution integral properly!

8.5.7 Correlation

The correlation process performs a measure of the similarity of two signals. It has several important practical applications. The most important is its ability to pick a known signal out of a sea of noise. The *cross-correlation* between $f(t)$ and $g(t)$ is defined as

$$\psi_{fg}(t) = \int_{-\infty}^{\infty} d\tau \, f(\tau)g(\tau - t). \tag{8.66}$$

The cross-correlation of a function with itself,

$$\psi_{ff}(t) = \int_{-\infty}^{\infty} d\tau \, f(\tau)f(\tau - t), \tag{8.67}$$
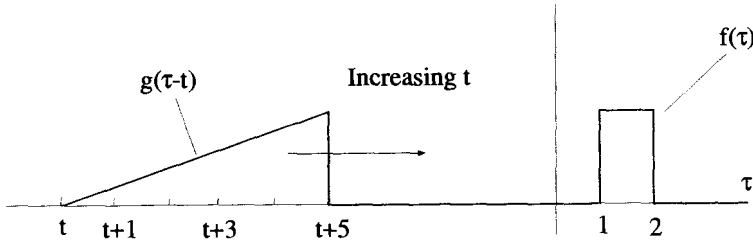
Figure 8.16   The Cross-Correlation Operation

is called an *autocorrelation*. It is frequently used to measure the pulse width of fast signals.

The correlation operation is quite similar to convolution described in the previous section, except the second function is not inverted. The t-variable shifts $g(\tau)$ with respect to $f(\tau)$ so that the functions of the integrand of the correlation operation appear as shown in Figure 8.16 if $f(t)$ and $g(t)$ are as defined in Figure 8.7. In general, the correlation operation requires the same care in breaking up the integral as is necessary for convolution.

The Fourier transform of the cross-correlation of two functions can be expressed in terms of the individual Fourier transforms of the two functions. The derivation follows the same steps used for convolution, and the result is

$$\mathcal{F}\left\{\int_{-\infty}^{\infty} d\tau\, f(\tau)g(\tau - t)\right\} = \sqrt{2\pi}\, \underline{F}(\omega)\underline{G}(-\omega). \tag{8.68}$$

8.5.8   Summary of Transform Properties

The properties of the Fourier transform are summarized in the table below. In this table a double arrow $\leftrightarrow$ is used to indicate the transform and its reverse process.

$$
\begin{array}{ccc}
f(t) & \leftrightarrow & \underline{F}(\omega) \\
g(t) & \leftrightarrow & \underline{G}(\omega) \\
f(t - t_o) & \leftrightarrow & e^{-i\omega t_o}\underline{F}(\omega) \\
df(t)/dt & \leftrightarrow & i\omega\underline{F}(\omega) \\
f(t)g(t) & \leftrightarrow & (1/\sqrt{2\pi})\underline{F}(\omega) \otimes \underline{G}(\omega) \\
f(t) \otimes g(t) & \leftrightarrow & \sqrt{2\pi}\underline{F}(\omega)\underline{G}(\omega) \\
\int_{-\infty}^{\infty} d\tau\, f(\tau)g(\tau - t) & \leftrightarrow & \sqrt{2\pi}\underline{F}(\omega)\underline{G}(-\omega) \\
f(-t) & \leftrightarrow & \underline{F}^{*}(\omega) \\
f(-t) = f(t) & \leftrightarrow & \underline{F}^{*}(\omega) = \underline{F}(\omega) \\
f(-t) = -f(t) & \leftrightarrow & \underline{F}^{*}(\omega) = -\underline{F}(\omega) \\
f(t) = f^{*}(t) & \leftrightarrow & \underline{F}(-\omega) = \underline{F}^{*}(\omega)
\end{array}
$$

The last four relations are true only if $f(t)$ is a pure real function of t. The others are valid even if $f(t)$ and $g(t)$ are complex functions.

8.6   FOURIER TRANSFORMS—EXAMPLES

This section derives the Fourier transforms of several common functions. In each case, we have attempted to explore an important technique or principle associated with the Fourier transform.

8.6.1   The Square Pulse

Consider the square pulse shown in Figure 8.17 with

$$f(t) = \begin{cases} 1 & -T/2 < t < T/2 \\ 0 & \text{otherwise} \end{cases}. \tag{8.69}$$

Clearly

$$\int_{-\infty}^{\infty} dt \; |f(t)| = T < \infty, \tag{8.70}$$

so this function is absolutely integrable and should have a valid Fourier transform. Using the definition of the transform,

$$\underline{F}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-T/2}^{T/2} dt \; e^{-i\omega t}. \tag{8.71}$$

This integration is straightforward and evaluates to

$$\underline{F}(\omega) = \left(\frac{1}{\sqrt{2\pi}} \right) \left(\frac{-e^{-i\omega t}}{i\omega} \right) \Big|_{-T/2}^{T/2}$$

$$= \sqrt{\frac{2}{\pi}} \left[\frac{\sin(\omega T/2)}{\omega} \right]. \tag{8.72}$$

This spectrum is plotted in Figure 8.18. It sits inside an envelope given by $\sqrt{2/(\pi\omega^2)}$. As $\omega \to 0$, it has the finite limit of $T/\sqrt{2\pi}$, because both the numerator and denominator in Equation 8.72 are going to zero at the same rate. The first zero of $\underline{F}(\omega)$ occurs when $\sin(\omega T/2) = 0$ or at $\omega = (2\pi)/T$. The other zeros are periodic at
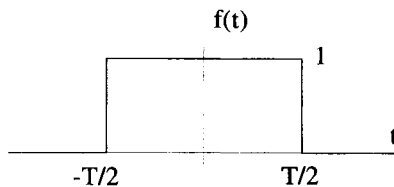
Figure 8.17   The Square Pulse
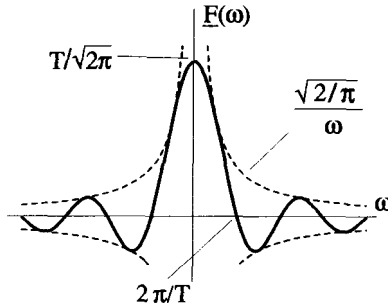
Figure 8.18 Spectrum of a Square Pulse

$\omega = 2\pi n/T$ where n is an integer. This function is so common that it has been given its own name:

$$\frac{\sin x}{x} \equiv \text{sinc } x. \tag{8.73}$$

We can easily find the area under any $\underline{F}(\omega)$ spectrum by evaluating the inverse Fourier transform of Equation 8.9 at $t = 0$:

$$\int_{-\infty}^{\infty} d\omega \, \underline{F}(\omega) = \sqrt{2\pi} f(0). \tag{8.74}$$

Therefore, for the square pulse

$$\int_{-\infty}^{\infty} d\omega \, \underline{F}(\omega) = \sqrt{2\pi}. \tag{8.75}$$

The area under the spectrum of the square pulse is finite and independent of T.

This has interesting consequences for the spectrum in the limit of $T \to \infty$. From Figure 8.18, notice that as T increases, the value of the spectrum at $\omega = 0$ increases, and all the zero crossings move toward $\omega = 0$. This happens in such a way that the area under the spectrum remains constant. In the limit of $T \to \infty$, therefore, this spectrum has all the qualities of a Dirac δ-function. This can be confirmed by combining the formal expression for $\underline{F}(\omega)$ with the orthogonality condition derived in Equation 8.17:

$$\underline{F}(\omega) = \lim_{T \to \infty} \frac{1}{\sqrt{2\pi}} \int_{-T/2}^{T/2} dt \, e^{-i\omega t}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \, e^{-i\omega t}$$

$$= \sqrt{2\pi} \delta(\omega). \tag{8.76}$$

Remember, however, that this is not a rigorous equation, because the δ-function is not inside an integral. The Fourier transform for the function $f(t) = 1$, is not really defined, since

$$\underline{F}(\omega) = \int_{-\infty}^{\infty} dt \, e^{-i\omega t} \tag{8.77}$$

does not converge. But in a casual sense, the δ-function in Equation 8.76 makes perfect sense, because when it is inserted into the inverse Fourier transform, we get back our original function $f(t) = 1$:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \, e^{i\omega t} \underline{F}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \, e^{i\omega t} \sqrt{2\pi} \delta(\omega)$$

$$= 1$$

We will explore the transform properties of the δ-function in greater detail in the following sections.

8.6.2 Transform of a δ-Function

In the light of the previous example, it is appropriate to ask for the Fourier transform of $\delta(t)$. Again, we will suspend the rigorous treatment of the δ-function and allow it to exist outside of integral operations. The δ-function is certainly absolutely integrable, because

$$\int_{-\infty}^{\infty} dt \, |f(t)| = \int_{-\infty}^{\infty} dt \, \delta(t) = 1, \tag{8.78}$$

and the transform itself is easy:

$$\underline{F}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \, e^{-i\omega t} \delta(t)$$

$$= \frac{1}{\sqrt{2\pi}}. \tag{8.79}$$

The inversion of this transform function becomes

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \, \frac{e^{i\omega t}}{\sqrt{2\pi}}, \tag{8.80}$$

which, using Equation 8.17, can be interpreted as $\delta(t)$.

The Fourier transform of a δ-function can be used to obtain the transform of a square pulse. If $f(t)$ is the square pulse shown in Figure 8.17, its derivative is the sum of two δ-functions

$$\frac{df(t)}{dt} = \delta(t + T/2) - \delta(t - T/2). \tag{8.81}$$

The transform of these two shifted δ-functions, using Equation 8.34, is

$$\mathcal{F}\{\delta(t + T/2) - \delta(t - T/2)\} = \frac{1}{\sqrt{2\pi}}[e^{i\omega T/2} - e^{-i\omega T/2}]. \qquad (8.82)$$

Because this is the transform of the derivative of the square pulse, Equation 8.46 says the spectrum of the square pulse is given by

$$\underline{F}(\omega) = \frac{1}{i\omega\sqrt{2\pi}}[e^{i\omega T/2} - e^{-i\omega T/2}]$$

$$= \sqrt{\frac{2}{\pi}}\left[\frac{\sin(\omega T/2)}{\omega}\right], \qquad (8.83)$$

which agrees with the result we found earlier in Equation 8.72.

8.6.3 Transform of a Gaussian

A normalized Gaussian pulse has the general form

$$f(t) = \frac{\alpha}{\sqrt{\pi}} e^{-\alpha^2 t^2}, \qquad (8.84)$$

where the leading constant normalizes the function, as we will prove shortly. A graph of Equation 8.84 is shown in Figure 8.19. Notice how the parameter α scales both the width and the height of the pulse.

The area of the pulse is given by

$$\text{Area} = \int_{-\infty}^{\infty} dt \, \frac{\alpha}{\sqrt{\pi}} e^{-\alpha^2 t^2} \qquad (8.85)$$
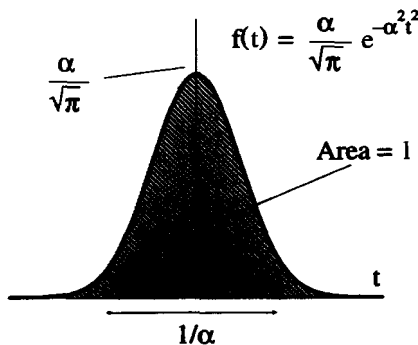
Figure 8.19   The Gaussian Pulse

and can be evaluated with a standard trick. First, make a substitution of $x = \alpha t$:

$$\text{Area} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} dx \, e^{-x^2}. \tag{8.86}$$

Now consider an equation for the square of the area:

$$\begin{aligned}
[\text{Area}]^2 &= \frac{1}{\pi} \int_{-\infty}^{\infty} dx \, e^{-x^2} \int_{-\infty}^{\infty} dy \, e^{-y^2} \\
&= \frac{1}{\pi} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \, e^{-(x^2 + y^2)}.
\end{aligned} \tag{8.87}$$

Notice the final integral of Equation 8.87 can be viewed as an integral over the entire Cartesian xy-plane. By changing the variables to a polar form with the substitutions $x^2 + y^2 = \rho^2$ and $dx \, dy = \rho \, d\rho \, d\theta$, Equation 8.87 becomes

$$[\text{Area}]^2 = \frac{1}{\pi} \int_0^{\infty} d\rho \int_0^{2\pi} d\theta \, \rho e^{-\rho^2}. \tag{8.88}$$

The $d\theta$ integration simply gives a factor of 2π, and the $d\rho$ integration that remains is straightforward. The final result is

$$[\text{Area}]^2 = 1. \tag{8.89}$$

Therefore,

$$\int_{-\infty}^{\infty} dt \, \frac{\alpha}{\sqrt{\pi}} \, e^{-\alpha^2 t^2} = 1. \tag{8.90}$$

The area under the normalized Gaussian is one, independent of the value of α.

The Fourier transform of the normalized Gaussian is

$$\begin{aligned}
\underline{F}(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \, \frac{\alpha}{\sqrt{\pi}} e^{-\alpha^2 t^2} e^{-i\omega t} \\
&= \frac{\alpha}{\pi\sqrt{2}} \int_{-\infty}^{\infty} dt \, e^{-(\alpha^2 t^2 + i\omega t)}.
\end{aligned} \tag{8.91}$$

This integration can be evaluated by *completing the square*. A quantity γ is added and subtracted from the argument of exponential function,

$$\left(\alpha^2 t^2 + i\omega t + \gamma \right) - \gamma, \tag{8.92}$$

so that the expression in the brackets of Equation 8.92 forms a perfect square. That is, we want to be able to write $\alpha^2 t^2 + i\omega t + \gamma$ as $(\alpha t + \beta)^2$. The value of β is therefore

determined by

$$\alpha^2 t^2 + i\omega t + \gamma = (\alpha t + \beta)^2$$
$$= \alpha^2 t^2 + 2\alpha\beta t + \beta^2. \tag{8.93}$$

From Equation 8.93,

$$2\alpha\beta t = i\omega t \tag{8.94}$$

so that

$$\beta = \frac{i\omega}{2\alpha} \tag{8.95}$$

and

$$\gamma = \beta^2 = -\frac{\omega^2}{4\alpha^2}. \tag{8.96}$$

Equation 8.91 can now be written as

$$\underline{F}(\omega) = \frac{\alpha}{\pi\sqrt{2}} e^{-\omega^2/(4\alpha^2)} \int_{-\infty}^{\infty} dt \; e^{-[\alpha t + i\omega/(2\alpha)]^2}. \tag{8.97}$$

This integration is accomplished using the substitution $x = \alpha t + i\omega/(2\alpha)$ and $dx = \alpha \, dt$:

$$\underline{F}(\omega) = \frac{1}{\pi\sqrt{2}} e^{-\omega^2/(4\alpha^2)} \int_{-\infty}^{\infty} dx \; e^{-x^2}$$
$$= \frac{1}{\sqrt{2\pi}} e^{-\omega^2/(4\alpha^2)}. \tag{8.98}$$

Thus we have shown the Fourier transform of a Gaussian is another Gaussian:

$$\mathcal{F}\left[\frac{\alpha}{\sqrt{\pi}} e^{-\alpha^2 t^2} \right] = \frac{1}{\sqrt{2\pi}} e^{-\omega^2/(4\alpha^2)}. \tag{8.99}$$

As α increases, the width of the Gaussian pulse decreases, while the width of its Gaussian spectrum increases, as shown in Figure 8.20. This demonstrates an important general property of the Fourier transform. The width of the function $f(t)$ is always inversely proportional to the width of the transform $\underline{F}(\omega)$. In the case of the Gaussian, the width of $f(t)$ is roughly

$$\Delta t \approx \frac{1}{\alpha}, \tag{8.100}$$

while the width of $\underline{F}(\omega)$ is approximately

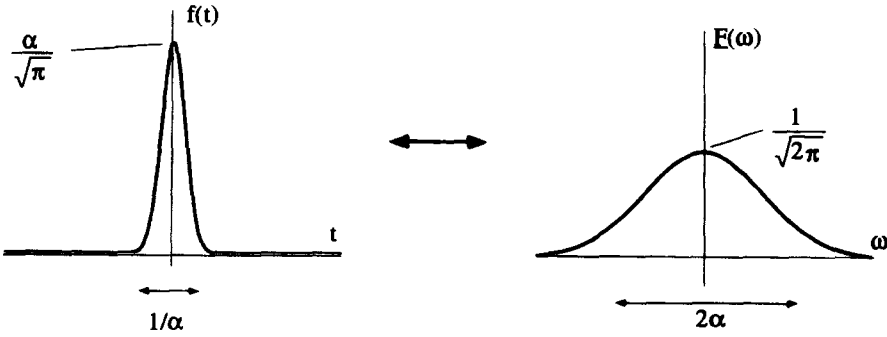$$\Delta\omega \approx 2\alpha. \tag{8.101}$$

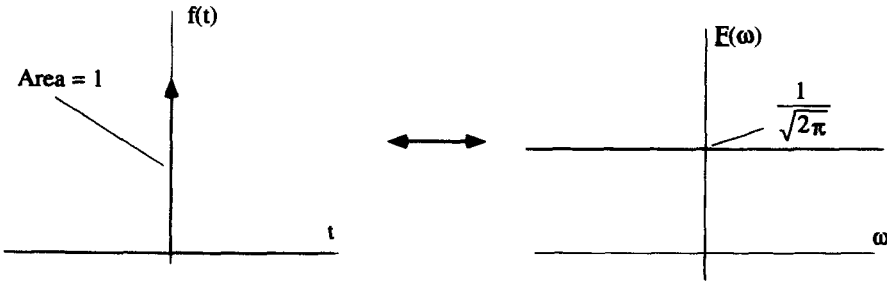**Figure 8.20**   The Gaussian and its Fourier Transform

Figure 8.21   The Gaussian and its Fourier Transform in the Limit $\alpha \rightarrow \infty$

Thus the product of the two is a constant:

$$\Delta \omega \Delta t \approx 2. \tag{8.102}$$

In general, for any functional shape, you will find that $\Delta \omega \Delta t \approx c$, where c is a constant determined by both the functional form and the precise definition of the "width" of a function. In the wave theory of quantum mechanics, this is the mathematical basis of the Heisenberg uncertainty principle.

The Gaussian can also be used to obtain the Fourier transform of the Dirac δ-function by looking at the limit of the Gaussian as $\alpha \rightarrow \infty$. In this limit, as can be seen from Figure 8.21, the Gaussian goes to $\delta(t)$ and its Fourier transform goes to $1/\sqrt{2\pi}$.

8.6.4   Periodic Functions

There is no formal Fourier transform for a periodic function, because the integral of Equation 8.20 cannot be performed. This limitation vanishes, if you are willing to allow Dirac δ-functions to appear outside of integral operations.

As an example, consider the function $f(t) = \sin t$. Its Fourier transform is given by

$$
\begin{aligned}
\underline{F}(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt\, e^{-i\omega t} \sin t \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt\, e^{-i\omega t} \left[\frac{1}{2i} \left(e^{it} - e^{-it} \right) \right] \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{2i} \int_{-\infty}^{\infty} dt\, \left[e^{it(-\omega+1)} - e^{it(-\omega-1)} \right] \\
&= i\sqrt{\frac{\pi}{2}} \left[\delta(\omega + 1) - \delta(\omega - 1) \right],
\end{aligned}
\tag{8.103}
$$

where we have made use of the informal orthogonality condition, Equation 8.18. Figure 8.22 shows this transform pair.

Fourier transforms which contain δ-functions, such as Equation 8.103, can be viewed as an alternative way of expressing the traditional Fourier series. Essentially, the δ-function converts the inverse Fourier integral into a Fourier series summation. For example, the inverse Fourier transform of the sine function is

$$
\begin{aligned}
f(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega\, e^{i\omega t} \left\{ i\sqrt{\frac{\pi}{2}} \left[\delta(\omega + 1) - \delta(\omega - 1) \right] \right\} \\
&= \frac{1}{2i} \left[e^{it} - e^{-it} \right] \\
&= \sum_{n=-\infty}^{\infty} \underline{c}_n e^{i\omega_n t},
\end{aligned}
\tag{8.104}
$$

where, for this function, all the $\underline{c}_n$ are zero except for $\underline{c}_1 = 1/(2i)$ and $\underline{c}_{-1} = -1/(2i)$. This function has a periodicity of $T_o = 2\pi$ and so $\omega_n = n$. This is simply the Fourier sum for the sine function.
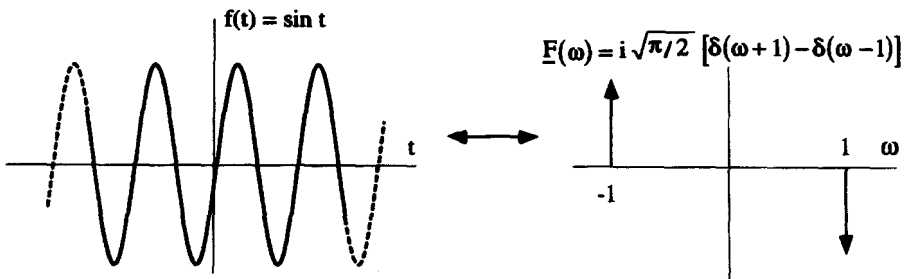

Figure 8.22   Fourier Transform of $\sin t$

Consequently, if you are willing to consider Dirac δ-functions outside of integral operations, the Fourier transform of periodic functions can be taken. In general, this transform will be a series of δ-functions which, when inserted into the inversion integral, generates the normal Fourier series. With this interpretation, a separate formalism for the Fourier series is unnecessary, because all the properties of the Fourier series are included in the Fourier transform!

8.6.5 An Infinite Train of δ-Functions

An interesting transform to consider is that of an infinite train of δ-functions:

$$f(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_o). \tag{8.105}$$

Because this is a periodic function, its Fourier transform must be composed of a sum of δ-functions. This is, however, a fairly difficult function to transform and must be done in steps.

We begin by taking the Fourier transform of a finite train of δ-functions:

$$f_{m_o}(t) = \sum_{n=-m_o}^{m_o} \delta(t - nT_o). \tag{8.106}$$

This function is absolutely integrable in the sense of Equation 8.23, and thus has a legitimate Fourier transform:

$$\begin{aligned}
\underline{F}_{m_o}(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt\, e^{-i\omega t} \sum_{n=-m_o}^{m_o} \delta(t - nT_o) \\
&= \frac{1}{\sqrt{2\pi}} \sum_{n=-m_o}^{m_o} \int_{-\infty}^{\infty} dt\, e^{-i\omega t} \delta(t - nT_o) \\
&= \frac{1}{\sqrt{2\pi}} \sum_{n=-m_o}^{m_o} e^{-i\omega nT_o}.
\end{aligned} \tag{8.107}$$

Because $f_{m_o}(t)$ is a real, even function of t, its transform is pure real for all values of ω.

The frequency dependence of this Fourier transform can be interpreted graphically. The transform is a sum of phasors, each of unit magnitude and at an angle of $-\omega nT_o$ in the complex plane. One such phasor is shown in Figure 8.23. With $\omega = 0$ all the phasors are at an angle of zero radians, as shown in Figure 8.24. From this figure, it is clear that the result of the sum is a single phasor given by

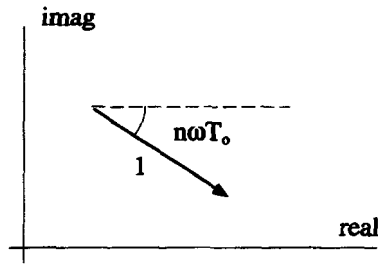$$\underline{F}_{m_o}(\omega = 0) = \frac{2m_o + 1}{\sqrt{2\pi}}. \tag{8.108}$$

Figure 8.23   One of the Phasors in the Fourier Transform of a Finite Train of δ-Functions
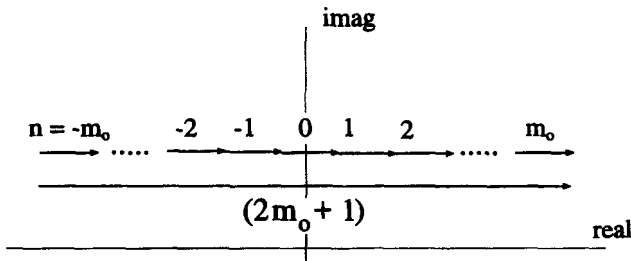

Figure 8.24   Fourier Transform for a Finite Train of δ-Functions Evaluated at $\omega = 0$

As ω increases from zero, the phasors in Figure 8.24 begin to curl up. The phasor for $n = 0$ is still at zero radians. The other phasors at $\pm n$ are at angles of $\mp \omega n T_o$. Because of the symmetry, the result is still a single phasor with a phase of zero radians, but its magnitude has decreased, as shown in Figure 8.25. When $\omega = 2\pi/T_o$, all the phasors are again aligned at zero radians so that $\underline{F}_{m_o}(\omega = 2\pi/T_o) = \underline{F}_{m_o}(\omega = 0)$. The Fourier transform repeats this maximum when ωT_o is any integer multiple of 2π.

A closed form expression for $\underline{F}_{m_o}(\omega)$, for arbitrary values of ω, can be obtained graphically. To see this consider a simplified version of the phasor sum

$$\sum_{n=-m_o}^{n=+m_o} e^{inx}. \tag{8.109}$$
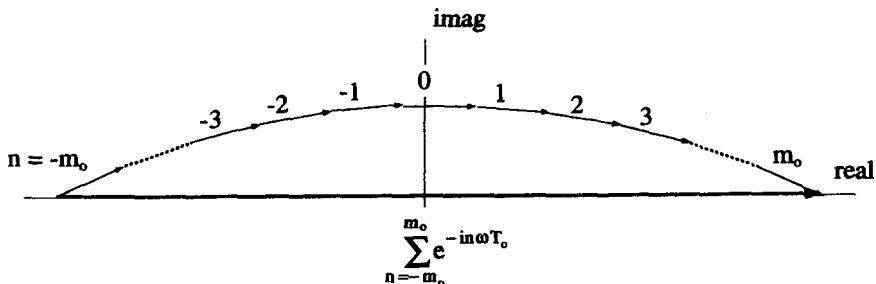
Figure 8.25   Fourier Transform for a Finite Train of δ-Functions as ω Increases from Zero
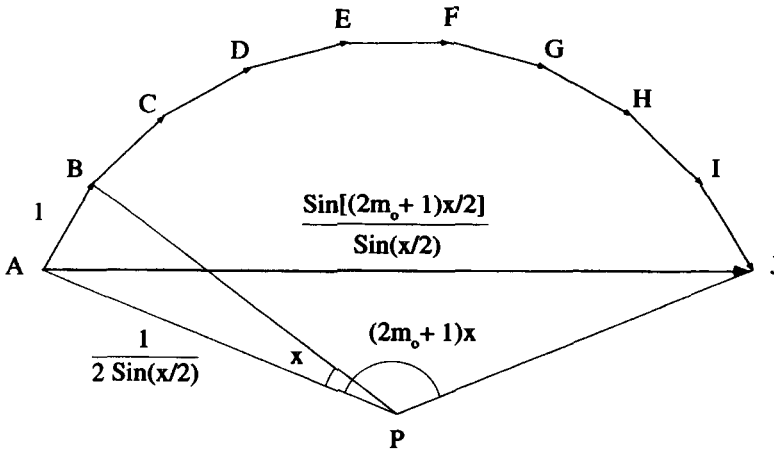
Figure 8.26 Phasor Summation

Equation 8.109 is a sum of $2m_o + 1$ phasors, each one of unit magnitude and at an angle of x radians relative to its predecessor in the complex plane. Figure 8.26 shows a geometric method for evaluating this sum. A point P is chosen which is equidistant from every intermediate point on the diagram. This distance is easy to determine, because the length of segment $\overline{AB}$ is one, and the angle $\angle APB$ is x. This makes $\overline{AP} = 1/[2\sin(x/2)]$. Then, because angle $\angle APJ = (2m_o + 1)x$, segment $\overline{AJ}$, which is the sum we seek, is given by

$$\overline{AJ} = \sum_{n=-m_o}^{m_o} e^{inx} = \frac{\sin\left[x(2m_o + 1)/2\right]}{\sin(x/2)}. \tag{8.110}$$

Putting the result of Equation 8.110 back into Equation 8.107 shows that the Fourier transform of a finite train of δ-functions is given by

$$\underline{F}_{m_o}(\omega) = \frac{1}{\sqrt{2\pi}} \frac{\sin[\omega T_o(2m_o + 1)/2]}{\sin(\omega T_o/2)}. \tag{8.111}$$

Figure 8.27 shows a graph of this function for $m_o = 3$.

We obtain the transform of an infinite train of δ-functions by taking the $m_o \to \infty$ limit of Equation 8.111. Notice, as m_o increases, the height of each peak in Figure 8.27 increases, while the distance between the peak value and the first zero shrinks. The distance between peaks remains a constant. The frequency of the oscillations in between these peaks increases without limit, so their effect cancels out inside of any integral, much like the sinc-sequence function described in Chapter 5.

It looks like we have all the characteristics of an infinite train of δ-functions! In order to validate this, we need to evaluate the integral area of one of the peaks. This
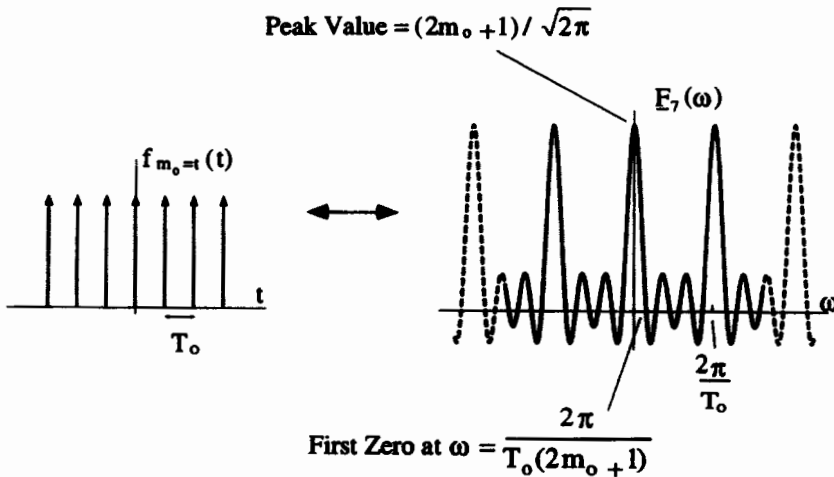
Peak Value $= (2m_o + 1)/\sqrt{2\pi}$

$\underline{F}_7(\omega)$

$f_{m_o = t}(t)$

T_o

t

$\frac{2\pi}{T_o}$

First Zero at $\omega = \dfrac{2\pi}{T_o(2m_o + 1)}$

ω

Figure 8.27   Fourier Transform for a Finite Number of δ-Functions

is given by

$$\text{Area} = \lim_{m_o \to \infty} \int_{-\pi/T_o}^{\pi/T_o} d\omega \; \frac{\sin[\omega T_o(2m_o + 1)/2]}{\sqrt{2\pi}\sin(\omega T_o/2)}. \qquad (8.112)$$

An easier method than grinding through the integral is to consider the triangular area shown in Figure 8.28. The magnitude of this area is $\sqrt{2\pi}/T_o$. As m_o increases, the area of the triangle is a better and better approximation to the area of the integral. In the limit, they are the same.

Thus, the Fourier transform of an infinite train of unit area δ-functions separated by $\Delta t = T_o$ is another infinite train of δ-functions of area $\sqrt{2\pi}/T_o$, separated by $\Delta \omega = 2\pi/T_o$:

$$\mathcal{F}\left\{ \sum_{n=-\infty}^{\infty} \delta(t - nT_o) \right\} = \frac{\sqrt{2\pi}}{T_o} \sum_{n=-\infty}^{\infty} \delta(\omega - 2\pi n/T_o). \qquad (8.113)$$

$\underline{F}_{m_o}(\omega)$

$(2m_o + 1)/\sqrt{2\pi}$

$\text{Area} = \dfrac{\sqrt{2\pi}}{T_o}$

ω

$\dfrac{4\pi}{T_o(2m_o + 1)}$
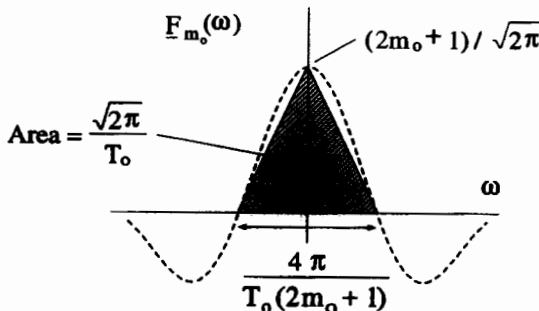
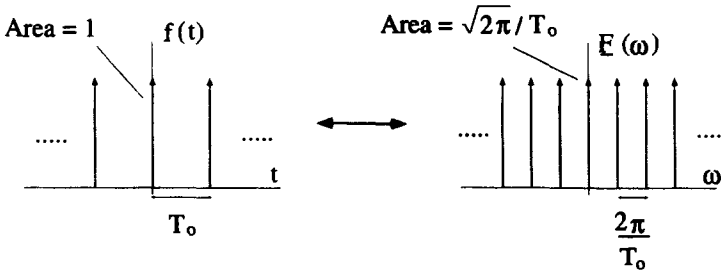**Figure 8.28**   Triangular Solution for the δ-Function Area m_o

Figure 8.29 Fourier Transform of an Infinite Train of δ-Functions

This is shown in Figure 8.29. Notice that as the δ-functions get closer together in t, the δ-functions of the spectrum move farther apart.

8.6.6 Transform of a Periodic Function Using Convolution

The transform of an infinite train of δ-functions can be coupled with convolution to find the Fourier transform of any periodic signal. Consider the function $h(t)$, which is a periodic train of square pulses, as shown in Figure 8.30. The pulse width is t_o, the pulse height is 1, and the distance between the start of one pulse and the start of the next pulse is T_o. Because $h(t)$ is a periodic function, its Fourier transform $\underline{H}(\omega)$ must be composed of δ-functions and can be obtained using a Fourier series integral. However, $h(t)$ can also be looked at as the convolution of two other functions

$$h(t) = f(t) \otimes g(t), \tag{8.114}$$

where $f(t)$ is a single square pulse of height 1 and width t_o, centered at $t = 0$, and $g(t)$ is an infinite train of unit area δ-functions separated by T_o, as shown in Figure 8.31.

Because $h(t)$ is the convolution of $f(t)$ with $g(t)$, its Fourier transform is just $\sqrt{2\pi}$ times the product of $\underline{F}(\omega)$ and $\underline{G}(\omega)$. These Fourier transforms have already been determined in previous examples:

$$\underline{F}(\omega) = \sqrt{\frac{2}{\pi}} \left[\frac{\sin(\omega t_o/2)}{\omega} \right] \tag{8.115}$$
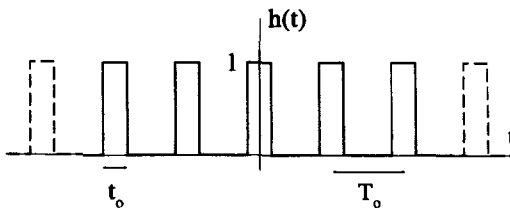

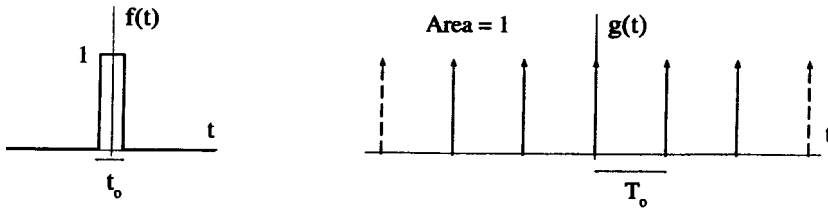**Figure 8.30** Periodic Train of Square Pulses

Figure 8.31   The Signals $f(t)$ and $g(t)$ which Convolve to Form $h(t)$

$$\underline{G}(\omega) = \frac{\sqrt{2\pi}}{T_o} \sum_{n=-\infty}^{\infty} \delta(\omega - 2\pi n/T_o). \qquad (8.116)$$

Consequently,

$$\underline{H}(\omega) = \frac{2\sqrt{2\pi}}{T_o} \left[\frac{\sin(\omega t_o/2)}{\omega} \right] \sum_{n=-\infty}^{\infty} \delta(\omega - 2\pi n/T_o). \qquad (8.117)$$

This is an infinite train of δ-functions with a δ-function located at every $\omega = 2\pi n/T_o$, each with an area weighted by the sinc function. If it is assumed that the pulse width is small compared to the period, i.e., $t_o \ll T_o$, the spectrum for this function looks as shown in Figure 8.32.

8.6.7   Periodic Burst

In the real world, there is no such thing as a periodic signal going on for all time. Real signals must start and stop, even if they look periodic over a long interval of time.
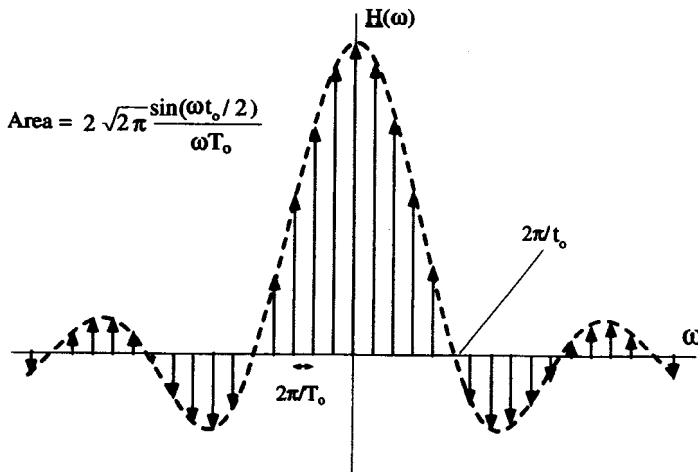

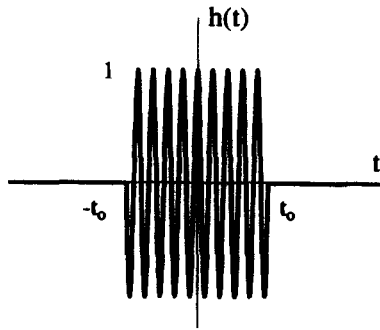**Figure 8.32**   Spectrum for a Train of Square Pulses

Figure 8.33   A $\cos(\omega_o t)$ Burst

The spectrum for signals of this type can be obtained using convolution techniques. Consider the function

$$h(t) = \begin{cases} \cos(\omega_o t) & -t_o < t < t_o \\ 0 & \text{otherwise} \end{cases} \qquad (8.118)$$

shown in Figure 8.33. We could calculate the Fourier transform of this signal using the standard Fourier integral. The transform is more easily obtained, however, by realizing that $h(t)$ is the product of two functions whose transforms are already known. That is, $h(t) = f(t)g(t)$, where $f(t)$ is a square pulse of width $2t_o$ and unit amplitude, and $g(t)$ is $\cos(\omega_o t)$. We already know the transforms of these two functions are

$$\underline{F}(\omega) = \sqrt{\frac{2}{\pi}} \frac{\sin(\omega t_o)}{\omega}$$

$$\underline{G}(\omega) = \sqrt{\frac{\pi}{2}} \left[\delta(\omega - \omega_o) + \delta(\omega + \omega_o) \right]. \qquad (8.119)$$

Because $h(t) = f(t)g(t)$,

$$\underline{H}(\omega) = \frac{1}{\sqrt{2\pi}} \underline{F}(\omega) \otimes \underline{G}(\omega)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega' \, \underline{F}(\omega')\underline{G}(\omega - \omega'). \qquad (8.120)$$

Figure 8.34 shows a diagram of this convolution operation.

In this convolution, $G(\omega - \omega')$ is the sum of two δ-functions. This makes the convolution integral relatively easy, because the δ-functions will just sift out the value of $F(\omega)$ where they overlap. If the expressions for $F(\omega)$ and $G(\omega)$ are inserted into Equation 8.120, the sifting property of the δ-functions results in

$$\underline{H}(\omega) = \frac{1}{\sqrt{2\pi}} \left\{ \frac{\sin[(\omega + \omega_o)t_o]}{\omega + \omega_o} + \frac{\sin[(\omega - \omega_o)t_o]}{\omega - \omega_o} \right\}, \qquad (8.121)$$

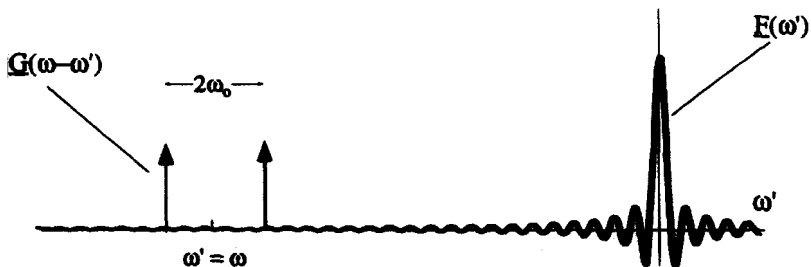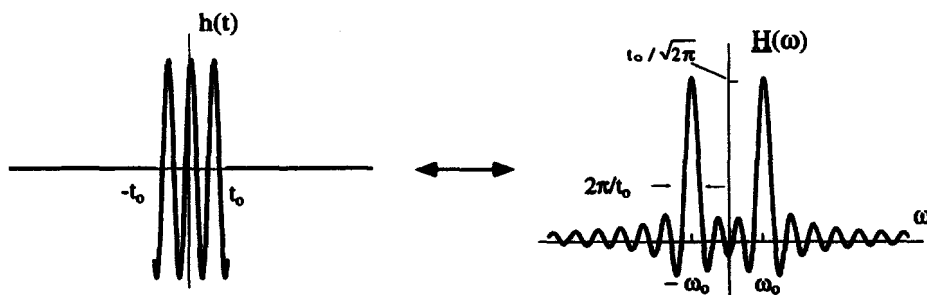**Figure 8.34**   Convolution Picture for the Burst Function

Figure 8.35   Transform of a Burst Signal

as shown in Figure 8.35. Notice how each δ-function creates a shifted version of $\underline{F}(\omega)$, and the two are summed to give $H(\omega)$. If ω_o is much larger than $2\pi/t_o$, the two copies of $\underline{F}(\omega)$ are separated and are essentially distinct. On the other hand, if ω_o is on the order of $2\pi/t_o$, there is significant overlap.

This transform has some interesting limits. One of these limits is when $t_o \to \infty$. In this case, the burst signal becomes more like an ideal periodic cosine signal. As you would expect, the pair of sinc functions in the transform become more like δ functions. Figure 8.36 depicts this limit. A second interesting limit is when $\omega_o \to 0$.

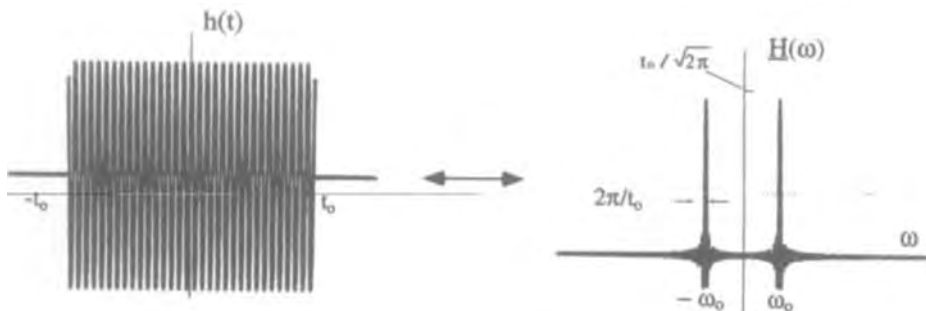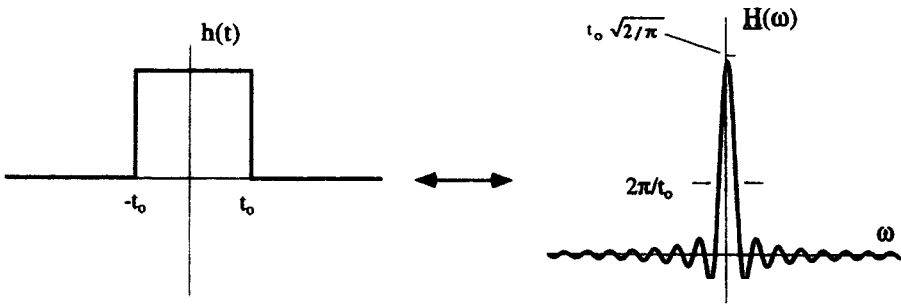**Figure 8.36**   Burst Transform as $t_o \to \infty$

Figure 8.37   Burst Transform as $\omega_o \rightarrow 0$

In this limit, $h(t)$ essentially becomes a square pulse, because $cos(\omega_o t) \rightarrow 1$. As you would expect from Equation 8.72, the transform approaches a single sinc function centered at $\omega = 0$. This is shown in Figure 8.37.

8.6.8   Exponential Decay

Consider the exponentially decaying function

$$f(t) = \begin{cases} 0 & t < 0 \\ e^{-\alpha_o t} & t > 0 \end{cases}, \tag{8.122}$$

where α_o is a real, positive number, as shown in Figure 8.38. This function is absolutely integrable, and its Fourier transform is easy to calculate:

$$\underline{F}(\omega) = \frac{1}{\sqrt{2\pi}} \int_0^\infty dt\, e^{-i\omega t} e^{-\alpha_o t}$$

$$= \frac{1}{i\sqrt{2\pi}} \left(\frac{1}{\omega - i\alpha_o} \right). \tag{8.123}$$

Notice that this transform has both a real and an imaginary part. This is because $f(t)$ has neither odd nor even symmetry.
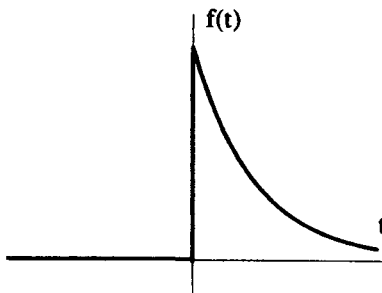
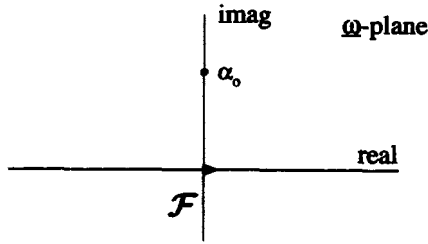

Figure 8.38   Exponential Decay

Figure 8.39   Inversion of Exponential Transform in the $\underline{\omega}$-Plane

Now consider the inverse transform back to $f(t)$:

$$
f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \, e^{i\omega t} \left[\frac{1}{i\sqrt{2\pi}} \left(\frac{1}{\omega - i\alpha_o} \right) \right]
$$

$$
= \frac{1}{2\pi i} \int_{-\infty}^{\infty} d\omega \, \frac{e^{i\omega t}}{\omega - i\alpha_o}. \tag{8.124}
$$

In Equation 8.124, ω is a pure real variable. The integral is easily calculated by extending the integral into the complex $\underline{\omega}$-plane, and then closing the contour. The complex integral is

$$
f(t) = \frac{1}{2\pi i} \int_{\mathcal{F}} d\underline{\omega} \, \frac{e^{i\underline{\omega} t}}{\underline{\omega} - i\alpha_o}, \tag{8.125}
$$

where the Fourier contour $\mathcal{F}$ is along the real axis, as shown in Figure 8.39. A single, first-order pole sits at $\underline{\omega} = i\alpha_o$.

If $t > 0$, we can get away with closing the contour using a semicircle of infinite radius even though the denominator of the integrand falls off only as fast as $1/\omega$. To justify this, we must show that the contribution along the path $\cap$, shown in Figure 8.40, vanishes as its radius goes to infinity. On $\cap$, $\underline{\omega} = R_o e^{i\theta}$ and $d\underline{\omega} = iR_o e^{i\theta} d\theta$, so

$$
\int_{\cap} d\underline{\omega} \, \frac{e^{i\underline{\omega} t}}{\underline{\omega} - i\alpha_o} = \lim_{R_o \to \infty} \int_0^{\pi} d\theta \, \frac{iR_o e^{i\theta} e^{iR_o t \, \cos\theta} e^{-R_o t \, \sin\theta}}{R_o e^{i\theta} - i\alpha_o}
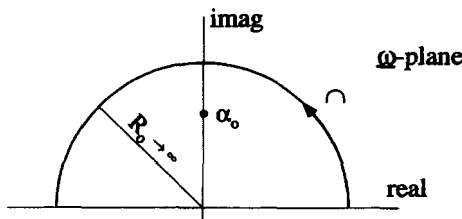$$

Figure 8.40   Upper Half-Plane Semicircle for Fourier Inversion

$$= \lim_{R_o \to \infty} \int_0^\pi d\theta \, \frac{iR_o e^{i\theta} e^{iR_o t \, \cos\theta} e^{-R_o t \, \sin\theta}}{R_o e^{i\theta}}$$

$$= \lim_{R_o \to \infty} \int_0^\pi d\theta \, ie^{iR_o t \, \cos\theta} e^{-R_o t \, \sin\theta}. \qquad (8.126)$$

Now consider the magnitude of the contribution from $\cap$:

$$\lim_{R_o \to \infty} \left| \int_\cap d\underline{\omega} \, \frac{e^{i\underline{\omega} t}}{\underline{\omega} - i\alpha_o} \right| = \lim_{R_o \to \infty} \left| \int_0^\pi d\theta \, ie^{iR_o t \, \cos\theta} e^{-R_o t \, \sin\theta} \right|$$

$$\leq \lim_{R_o \to \infty} \int_0^\pi d\theta \, e^{-R_o t \, \sin\theta}$$

$$\leq \lim_{R_o \to \infty} 2 \int_0^{\pi/2} d\theta \, e^{-R_o t \, \sin\theta}. \qquad (8.127)$$

The second step in this sequence comes from the fact that $\left| ie^{iR_o t \cos\theta} \right| = 1$, while the last step is true because the function is symmetric about $\theta = \pi/2$.

There is an important result, called Jordan's inequality, which can be used at this point. The inequality states that, for $t > 0$,

$$\int_0^{\pi/2} d\theta \, e^{-R_o t \, \sin\theta} \leq \frac{\pi}{2R_o t}. \qquad (8.128)$$

This is proved quite easily by noticing that $\sin\theta \geq 2\theta/\pi$ for $0 \leq \theta \leq \pi/2$. Using this result in Equation 8.127 gives

$$\lim_{R_o \to \infty} \left| \int_\cap d\underline{\omega} \, \frac{e^{i\underline{\omega} t}}{\underline{\omega} - i\alpha_o} \right| \leq \lim_{R_o \to \infty} \frac{\pi}{R_o t}$$

$$= 0. \qquad (8.129)$$

This shows that for $t > 0$, the contribution from $\cap$ vanishes as the radius grows infinite, as long as $\underline{F}(\underline{\omega})$ falls off as $1/|\omega|$ or faster.

For $t < 0$, we cannot close the contour in the upper plane as before because the integrand diverges for large, positive imaginary numbers. Instead, we must change the contour to close in the lower half plane. By arguments similar to the ones given above, you can show there is no contribution added when this contour is closed. That is,

$$\int_\cup d\underline{\omega} \, \frac{e^{i\underline{\omega} t}}{\underline{\omega} - i\alpha_o} = 0 \qquad t < 0, \qquad (8.130)$$

where the contour $\cup$ is a semicircle of infinite radius in the lower half $\underline{\omega}$-plane. Consequently, the integration of Equation 8.125 can be performed by closure

$$f(t) = \frac{1}{2\pi i} \oint_C d\underline{\omega} \frac{e^{i\underline{\omega} t}}{\underline{\omega} - i\alpha_o}, \qquad (8.131)$$
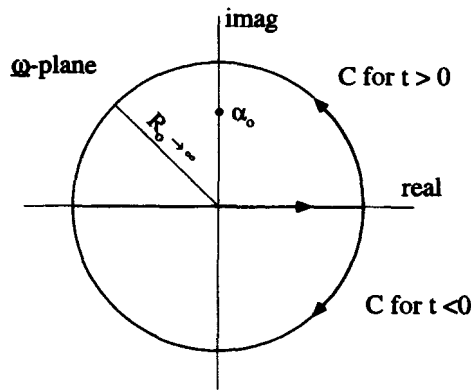
Figure 8.41   Closure of the Fourier Inversion Integration

where the contour C is closed in the upper half ω-plane for $t > 0$, and in the lower half ω-plane for $t < 0$, as shown in Figure 8.41.

For each contour, the integral can be determined by adding up the residues of the enclosed poles. The $t < 0$ contour has no enclosed singularities, and the $t > 0$ contour has just one, with a residue of $e^{-\alpha_o t}$. Thus our final result is

$$f(t) = \begin{cases} 0 & t < 0 \\ e^{-\alpha_o t} & t > 0 \end{cases}, \tag{8.132}$$

which is our original function.

Now, consider what happens as α_o approaches zero. This makes the $f(t)$ look more and more like a step function, as shown in Figure 8.42. For any α_o bigger than zero, we can obtain a valid Fourier transform, but if $\alpha_o = 0$, we know we cannot take the Fourier transform, because the integral

$$F(\omega) = \frac{1}{2\pi} \int_0^\infty dt \, e^{-i\omega t} \tag{8.133}$$
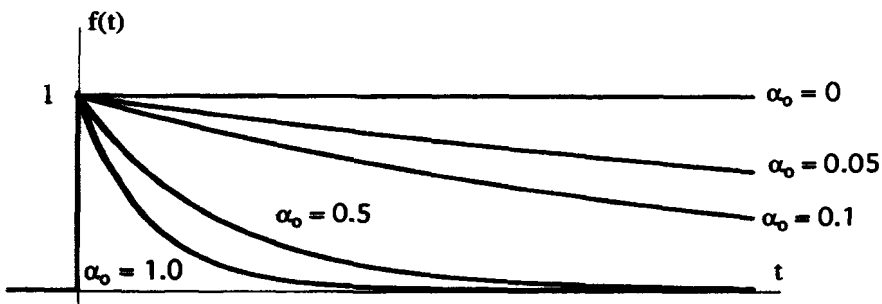
does not exist.

Figure 8.42   The Exponential Function in the Limit of $\alpha_o \to 0$
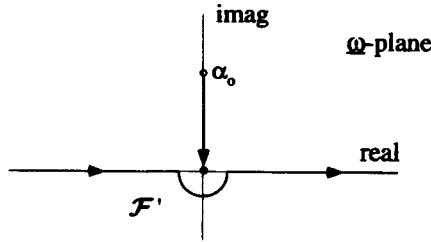
Figure 8.43 Modified Fourier Contour to Handle Exponential Function in the Limit of $\alpha_o \rightarrow 0$

By looking at what happens to the inversion contour for this function in the $\alpha_o \rightarrow 0$ limit, we can gain a valuable alternative viewpoint. As long as α_o is greater than zero, no matter how small, the pole of the integrand of Equation 8.125 is in the upper half $\underline{\omega}$-plane, and the $\mathcal{F}$ contour can remain on the real $\underline{\omega}$-axis. But when $\alpha_o = 0$, the pole sits right on top of the $\mathcal{F}$-contour and the integral cannot be performed.

Notice, however, that if the $\mathcal{F}$-contour is deformed to $\mathcal{F}'$ by dodging the pole, as shown in Figure 8.43, the inversion integral

$$f(t) = \frac{1}{2\pi i} \int_{\mathcal{F}'} d\underline{\omega} \, \frac{e^{i\underline{\omega}t}}{\underline{\omega}} \tag{8.134}$$

reproduces the unit step function

$$f(t) = \begin{cases} 0 & t < 0 \\ 1 & t > 0 \end{cases}. \tag{8.135}$$

This is no coincidence, and it will lead us to a more general integral transform called the Laplace transform.

8.6.9 Damped Sinusoid

As a last example, consider the damped sinusoid

$$f(t) = \begin{cases} 0 & t < 0 \\ e^{-\alpha_o t} \sin(\omega_o t) & t > 0 \end{cases}, \tag{8.136}$$

where α_o and ω_o are both positive numbers. Figure 8.44 shows this function. The Fourier transform is easy to calculate:

$$\underline{F}(\omega) = \frac{1}{\sqrt{2\pi}} \int_0^\infty dt \, e^{-i\omega t} e^{-\alpha_o t} \left(\frac{e^{i\omega_o t} - e^{-i\omega_o t}}{2i} \right)$$

$$= -\frac{\omega_o/\sqrt{2\pi}}{(\omega - \underline{\omega}_1)(\omega - \underline{\omega}_2)}, \tag{8.137}$$
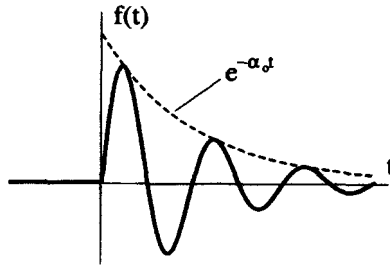
Figure 8.44   The Damped Sinusoid

where $\underline{\omega}_1 = \omega_o + i\alpha_o$ and $\underline{\omega}_2 = -\omega_o + i\alpha_o$. Notice $\underline{F}(\omega)$ is a complex function because $f(t)$ has neither odd nor even symmetry.

The inversion transform of Equation 8.137 is

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \, e^{i\omega t} \frac{-\omega_o/\sqrt{2\pi}}{(\omega - \underline{\omega}_1)(\omega - \underline{\omega}_2)}. \tag{8.138}$$

As with the exponential decay function of the previous section, the best way to evaluate this integral is to extend it into the complex plane, and close the complex contour. The complex version of the inversion integral is

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathcal{F}} d\underline{\omega} \, e^{i\underline{\omega}t} \frac{-\omega_o/\sqrt{2\pi}}{(\underline{\omega} - \underline{\omega}_1)(\underline{\omega} - \underline{\omega}_2)}, \tag{8.139}$$

where $\mathcal{F}$ is the Fourier contour along the real $\underline{\omega}$-axis, as shown in Figure 8.45. Also, like the exponential decay example, we are forced to close the contour in the upper half-plane for $t > 0$ and the lower half-plane for $t < 0$. This is due to the influence of the $e^{i\underline{\omega}t}$ term in the numerator. As you will prove in one of the exercises at the end of the chapter, Equation 8.139 returns the original function $f(t)$,

$$f(t) = \begin{cases} 0 & t < 0 \\ e^{-\alpha_o t} \sin(\omega_o t) & t > 0 \end{cases}, \tag{8.140}$$
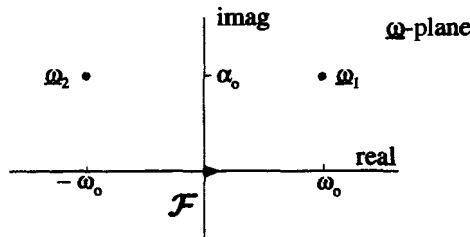
as long as $\alpha_o > 0$.

Figure 8.45   Inversion of the Damped Sinusoid Transform in the $\underline{\omega}$-Plane
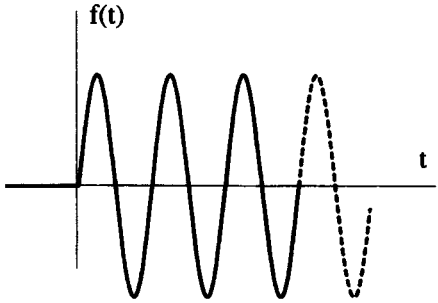
Figure 8.46   The Undamped Sinusoid

The Fourier transform and inversion for the damped sinusoid work correctly for all values of $\alpha_o > 0$, no matter how close to zero α_o gets. Exactly at $\alpha_o = 0$, however, we know we cannot have a valid Fourier transform, because now $f(t)$ becomes

$$f(t) = \begin{cases} 0 & t < 0 \\ \sin(\omega_o t) & t > 0 \end{cases} \tag{8.141}$$

as shown in Figure 8.46. The sinusoid is no longer damped, and the Fourier integral does not exist. This problem is also evident from looking at the inversion contour. With $\alpha_o = 0$, the Fourier transform in Equation 8.137 has become

$$\underline{F}(\omega) = -\frac{\omega_o/\sqrt{2\pi}}{(\omega - \omega_o)(\omega + \omega_o)}, \tag{8.142}$$

and the two poles in Figure 8.44 now lie on the Fourier contour.

Notice, however, Equation 8.142 still can be used to generate the undamped sinusoid

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathcal{F}'} d\underline{\omega}\, e^{i\underline{\omega}t} \frac{-\omega_o/\sqrt{2\pi}}{(\underline{\omega} - \omega_o)(\underline{\omega} + \omega_o)}, \tag{8.143}$$

if the Fourier contour is modified to $\mathcal{F}'$, as shown in Figure 8.47. In fact, even an exponentially growing sinusoid,
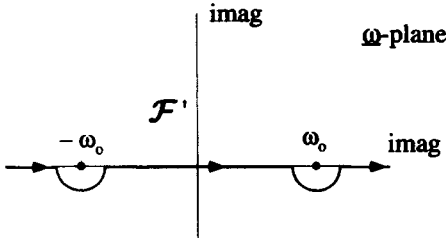

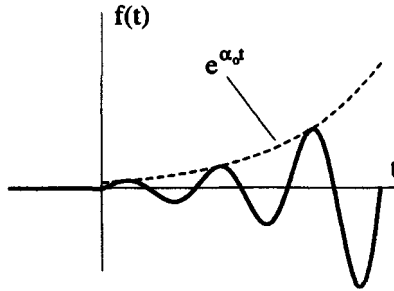**Figure 8.47**   Modified Contour to Handle an Undamped Sinusoid
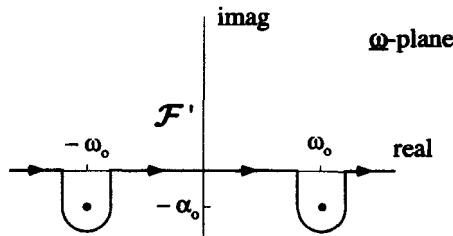
Figure 8.48   Exponentially Growing Sinusoid

Figure 8.49   Modified Contour for an Exponentially Growing Sinusoid

$$f(t) = \begin{cases} 0 & t < 0 \\ e^{\alpha_o t} \sin(\omega_o t) & t > 0 \end{cases}, \qquad (8.144)$$

such as the one shown in Figure 8.48, can be recovered from this inversion, if we let the $\mathcal{F}'$ contour dip below the poles of $\underline{F}(\omega)$, as shown in Figure 8.49.

These deformed contours seem to extend the Fourier inversion into realms where you would not expect it to work. In the next chapter we formally introduce these extensions as the basis for the Laplace transform.

8.7   THE SAMPLING THEOREM

One of the more important results of Fourier transform theory is the *Sampling Theorem*. This theorem determines at what rate a signal must be sampled so enough information is available to properly reproduce it. This is important, because sampling too infrequently will cause the reproduced signal quality to degrade, while requiring a sampling rate which is too high is inefficient.

Consider sampling an audio signal for recording onto a compact disc. Let the original microphone signal $a(t)$ look as shown in Figure 8.50. Furthermore, let this analog signal have a Fourier transform $\underline{A}(\omega)$, as shown in Figure 8.51. This spectrum vanishes above and below some cutoff frequency ω_{max}, which is perhaps imposed by
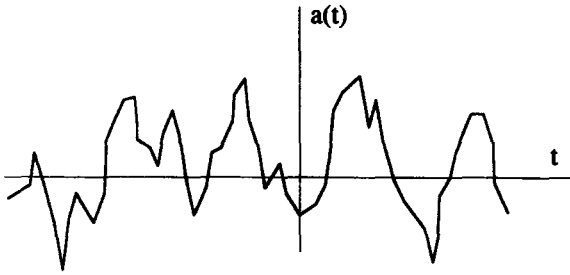
Figure 8.50 The Original Analog Audio Signal

the mechanics of the microphone. For the purposes of this discussion, the transform will be assumed to be pure real. This assumption is not necessary for the validity of the Sampling Theorem, but is convenient for our drawings.

A useful way to model the sampling process, and the signal recorded on the CD, is to let the sampled signal, which we will call $a_s(t)$, be equal to the original signal times a periodic set of δ-functions

$$a_s(t) = a(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_o). \tag{8.145}$$

The time between the δ-functions is the sampling interval. Figure 8.52 shows this representation.

Because $a_s(t)$ is the product of two functions, the spectrum of $a_s(t)$, which we will call $A_s(\omega)$, is the convolution of $A(\omega)$ with the transform of the train of δ-functions:

$$A_s(\omega) = \frac{1}{\sqrt{2\pi}} A(\omega) \otimes \frac{\sqrt{2\pi}}{T_o} \sum_{-\infty}^{\infty} \delta(\omega - 2\pi n/T_o). \tag{8.146}$$

When the sampling rate is very high, that is when $T_o \ll 2\pi/\omega_{max}$, $A_s(\omega)$ has the form shown in Figure 8.53. The spectrum of $a_s(t)$ is not the same as the spectrum of $a(t)$, because it has many more high frequency components. If it were reproduced through
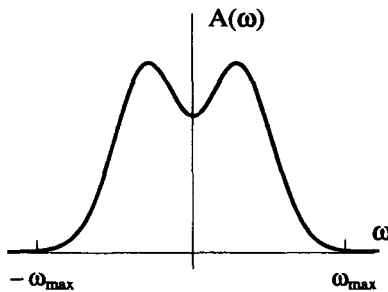
Figure 8.51 Fourier Transform of the Original Signal
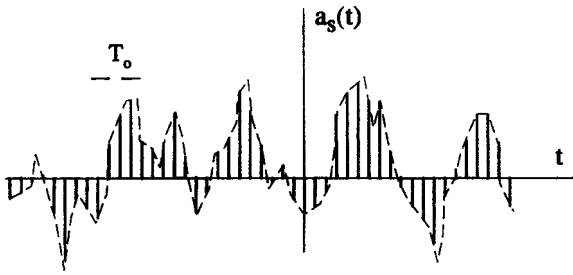
Figure 8.52   The Sampled Signal

an amplifier and speaker system, it would not sound like $a(t)$. However, sitting in the spectrum of $a_s(t)$, from $-\omega_{max}$ to $+\omega_{max}$, is the exact spectrum of $a(t)$. Thus, if we send this signal through a low-pass filter before amplification, the result should be a sound that reproduces $a(t)$.

A simple low-pass filter can be made from the capacitor and resistor circuit shown in Figure 8.54. The input voltage $v_{in}(t)$ and output voltage $v_{out}(t)$ are related by the differential equation

$$v_{in}(t) = RC\frac{dv_{out}(t)}{dt} + v_{out}(t).$$ (8.147)

Applying a Fourier transform to both sides of this equation gives an algebraic equation relating the Fourier transforms of the input and output waveforms:

$$\underline{V}_{in}(\omega) = (i\omega RC + 1)\underline{V}_{out}(\omega)$$ (8.148)

or

$$\underline{V}_{out}(\omega)/\underline{V}_{in}(\omega) = \frac{1}{1 + i\omega RC}.$$ (8.149)
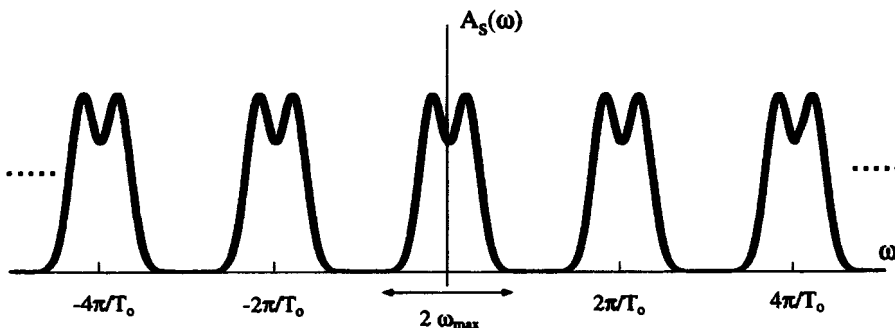
Figure 8.53   Spectrum of Sampled Signal
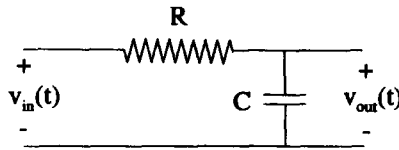
Figure 8.54   Simple Low-Pass Filter

The quantity $\underline{V}_{out}(\omega)/\underline{V}_{in}(\omega)$ is called the transfer function for the filter. The magnitude of this function is shown in Figure 8.55. For frequencies much less than $\omega = 1/RC$, the output voltage and input voltage are essentially identical. For frequencies much above $\omega = 1/RC$, the output is zero. This transition frequency is called the cutoff frequency of the filter, ω_{co}. For a simple RC filter, such as this one, the transition at the cutoff is not very sharp. Better filters can be made with a more complicated combination of components, or with more sophisticated digital processing components. The ideal low-pass filter has the transfer function shown in Figure 8.56.

If we pass the sampled signal through an ideal low-pass filter with $\omega_{co} = \omega_{max}$, the output of the filter will have a Fourier transform $\underline{A}_f(\omega)$ that is exactly equal to $\underline{A}(\omega)$, and consequently the signal coming out of the filter $a_f(t)$ will sound exactly like the original $a(t)$ signal from the microphone. The filter has removed the spurious high frequency components introduced by the sampling. This process is shown in

Figure 8.55   Transfer Function for a Simple RC Low-Pass Filter
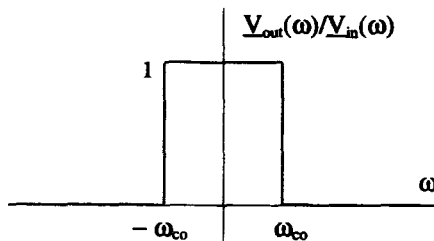
Figure 8.56   Ideal Low-Pass Filter
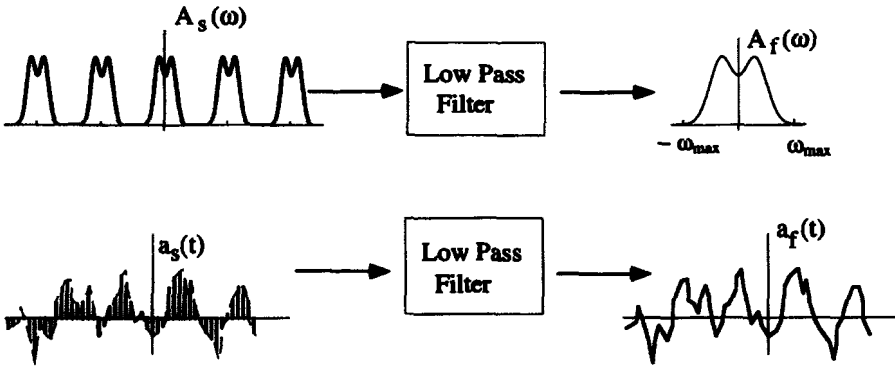
Figure 8.57   Frequency and Time Domain Pictures of the Filtering Process

Figure 8.57. The top picture in Figure 8.57 shows the filtering process in the frequency domain, while the bottom picture shows the same process in the time domain.

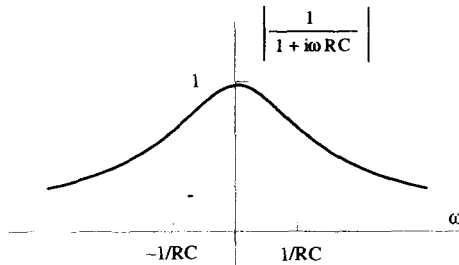In the previous arguments, we made the important assumption that the original signal was sampled at a high rate. The spectrum for $\underline{A}_s(\omega)$ is composed of a series of $\underline{A}(\omega)$ spectra replicated every $\omega = 2n\pi/T_o$. As long as $\pi/T_o > \omega_{max}$, these spectrum do not overlap, and our previous analysis works fine. But when T_o grows so large that $\pi/T_o < \omega_{max}$, the individual spectra overlap each other and $\underline{A}_s(\omega)$ looks as shown in Figure 8.58. Now, the output of the low-pass filter with $\omega_{co} = \omega_{max}$ gives the spectrum shown in Figure 8.59. In this case the signal coming out of the filter would no longer sound like the signal from the microphone.

Thus we see, in order to preserve the information in a signal, it must be sampled such that $\pi/T_o > \omega_{max}$. This means the sampling frequency $2\pi/T_o$ must be at least twice the highest frequency present in the sampled signal. The minimum sampling frequency is often called the Nyquist sampling rate.

Keep in mind that we have drastically simplified the process of CD sampling in our previous argument. Some practical issues that introduce complications should be pointed out. First, in real CD sampling, the recorded audio signal is digitized
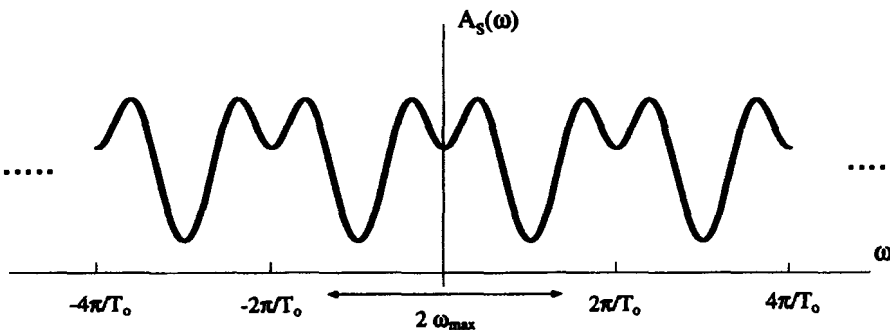
Figure 8.58   Spectrum of an Undersampled Signal
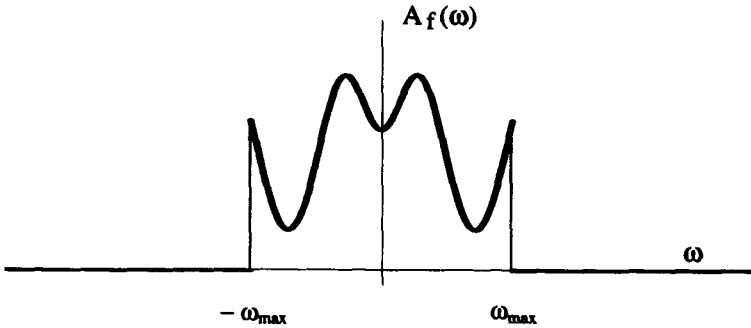
Figure 8.59 Spectrum of the Filtered, Undersampled Signal

and can take on only discrete values. This requires the use of an analog-to-digital converter (ADC). The recreation of the original signal then requires the inverse step of a digital-to-analog conversion (DAC). Second, for an audio signal, the practical limit for ω_{max} is about 20 kHz, the frequency limit of the human ear. The music industry has settled on a 44 Khz sampling standard, just above twice the maximum frequency we can hear. However, the microphone signal and its electronic handling before sampling may very well have frequency components well above 22 kHz. To prevent the overlap in the convolved spectra for $\underline{A}_s(\omega)$ that would result by sampling such a signal, a low-pass filter with a cutoff frequency near 22 kHz is inserted just before the sampling electronics. Finally, there is no such thing as an ideal filter with a transfer characteristic like that shown in Figure 8.56; there will always be some "rolloff." The better the filter, the better the sound reproduction.

EXERCISES FOR CHAPTER 8

1. What is the Fourier Transform of the function

$$f(t) = \begin{cases} e^{-\alpha t} + e^{-2\alpha t} & t > 0 \\ 0 & t < 0 \end{cases},$$

where α is real and positive?

2. Find the Fourier transform of the function

$$f(t) = \begin{cases} \sin t & -\pi/2 < t < \pi/2 \\ 0 & \text{otherwise} \end{cases}.$$

Is there a way to evaluate this without actually calculating the transform integral?

3. Find the Fourier transform of the signal shown below,
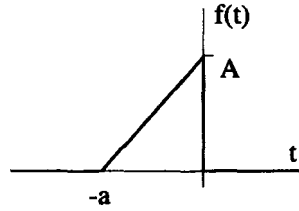
(a) by directly calculating the transform integral.

(b) by differentiating the signal and writing the transform in terms of transforms that were already derived in this chapter.

4. In this chapter, we determined the Fourier transform of the Dirac δ-function. Find the Fourier transform of the derivative of the Dirac δ-function,

$$f(t) = \frac{d\delta(t)}{dt}.$$

Compare these two transforms.

5. Find the Fourier transform of the function

$$f(t) = \begin{cases} t^3 & 0 < t < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) by directly calculating the transform integral using integration by parts.

(b) by differentiating the signal three times and writing the transform in terms of the sum of two simpler transforms.

6. The Fourier transform of $f(t)$ is

$$\underline{F}(\omega) = \frac{-i}{\omega^2 - i2\alpha\omega - \alpha^2}.$$

Determine the function $f(t)$ and make a plot of it for positive and negative values of t. In the limit of $\alpha \to 0$, what happens to $\underline{F}(\omega)$ and $f(t)$?

7. Obtain the Fourier transform of the function

$$f(t) = \begin{cases} e^{-\alpha_o(t-\tau)} \sin{[\omega(t-\tau)]} & t > \tau \\ 0 & t < \tau \end{cases}.$$

Show that the inversion of this Fourier transform requires a slight modification to the standard $t > 0$, $t < 0$ closure semicircles discussed in this chapter.
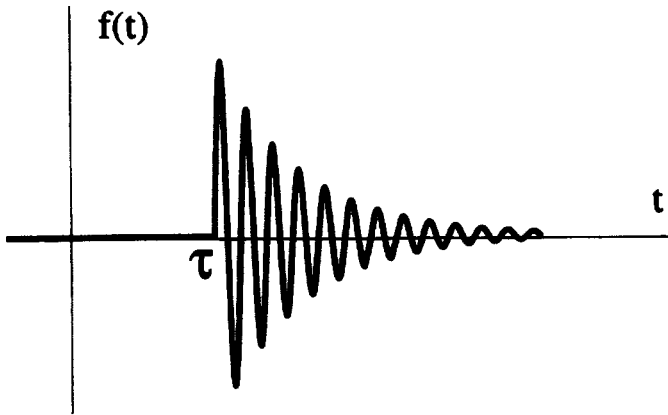
8. Consider the Fourier inversion

$$f(x) = \int_{-\infty}^{\infty} dk \, \frac{e^{ikx}}{k - i},$$

where x and k are pure real variables.

(a) For $x > 0$, determine how this integral can be closed in the complex $\underline{k}$-plane and evaluate $f(x)$.

(b) For $x < 0$, determine how this integral can be closed in the complex $\underline{k}$-plane and evaluate $f(x)$.

9. Prove that the inversion of Equation 8.139 gives Equation 8.140.

10. Given the functions $f(t)$ and $g(t)$ shown below,
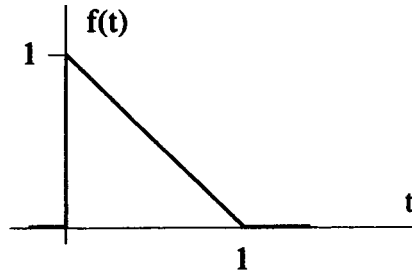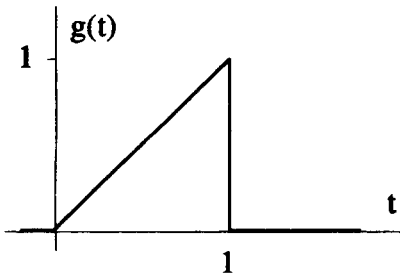
(a) Make a labeled sketch of $f(t) \otimes g(t)$.

(b) Make a labeled sketch of $g(t) \otimes g(t)$.

(c) Make a labeled sketch of the autocorrelation of $g(t)$.

11. Consider the two functions

$$f(t) = \begin{cases} 1/2 & -1 < t < +1 \\ 0 & \text{otherwise} \end{cases}$$

$$g(t) = t\, e^{-t^2}.$$

(a) Plot $f(t)$ and $g(t)$ for all t.

(b) Find an expression for $h(t)$ where $h(t) = f(t) \otimes g(t)$.

(c) Plot $h(t)$.

12. Consider the two functions:

$$f(t) = \begin{cases} t^2 & -1 < t < +1 \\ 0 & \text{otherwise} \end{cases}$$

$$g(t) = \begin{cases} 1 & 0 < t < +1 \\ 0 & \text{otherwise} \end{cases}.$$

Make a labeled sketch of $f(t)$ convolved with $g(t)$.

13. Let $f(t)$ be a square pulse

$$f(t) = \begin{cases} 1 & -T_o/2 < t < +T_o/2 \\ 0 & \text{otherwise} \end{cases}.$$

and $g(t)$ be made up of a delayed version of $f(t)$ plus a sinusoidal function of time:

$$g(t) = f(t - 10T_o) + \sin(\omega_o t),$$

where $\omega_o = 20\pi/T_o$. Make a labeled sketch of the cross-correlation of $f(t)$ with $g(t)$.

14. Show that the convolution integral is symmetric. That is, prove

$$\int_{-\infty}^{\infty} d\tau\, f(\tau) g(t - \tau) = \int_{-\infty}^{\infty} d\tau\, g(\tau) f(t - \tau).$$

15. Show that if both $g(t)$ and $f(t)$ are zero for $t < 0$, then the convolution of the two functions $g(t) \otimes f(t)$ is also zero for $t < 0$.

16. Find the Fourier transform of the cross-correlation of $f(t)$ with $g(t)$ in terms of the Fourier transforms of the individual transforms $\underline{F}(\omega)$ and $\underline{G}(\omega)$.

17. Use the fact that the Fourier transform of a Gaussian is a Gaussian,

$$\frac{\alpha}{\sqrt{\pi}} e^{-\alpha^2 t^2} \leftrightarrow \frac{1}{\sqrt{2\pi}} e^{-\omega^2/(4\alpha^2)},$$

to evaluate the integral

$$h(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\tau\, e^{-\tau^2}\, e^{-(t-\tau)^2}.$$

On a single graph plot $h(t)$ and e^{-t^2}.

18. The function $f(t)$ is a periodic train of Gaussian pulses with period T, as described below.

$$f(t) = \frac{\alpha}{\sqrt{\pi}} \sum_{n=-\infty}^{\infty} e^{-\alpha^2(t-nT)^2}$$
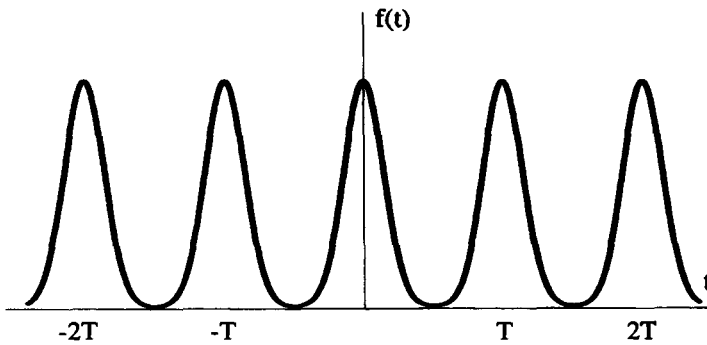
This function can be looked at as the convolution of a single Gaussian pulse with a train of Dirac δ-functions. Using this, determine the Fourier transform of $f(t)$.

19. Find the Fourier transform for the periodic function $f(t)$ shown below.
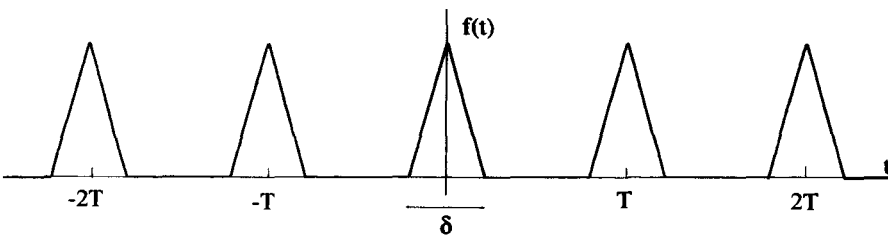
Look at $f(t)$ as the convolution of two functions.

20. Find and plot the Fourier transform for the periodic function $f(t)$ where

$$f(t) = \begin{cases} e^{-\alpha_o t} \sin t & 0 < t < 6\pi \\ 0 & 6\pi < t < 60\pi \\ f(t + 60\pi) & \text{all } t \end{cases}.$$
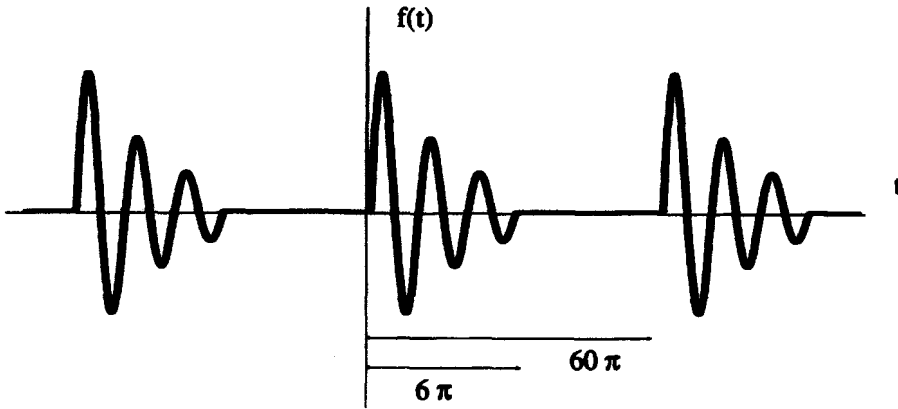
21. To evaluate real definite integrals, we have often used the technique of closure in the complex plane. We used a similar process to invert several of the complex Fourier transforms developed in this chapter. To be able to use closure, the functions being integrated must meet certain conditions. Show that the conditions for closure are different for a real, definite integral than they are for a Fourier inversion.

For this problem, consider the function:

$$F(\omega) = \frac{\omega^2}{\omega^3 + 1}.$$

(a) Show that, if the integral

$$\int_{-\infty}^{\infty} d\omega\, F(\omega)$$

is extended into the complex ω-plane, closure is not possible.

(b) Show that the inverse transform

$$f(t) = \int_{\infty}^{\infty} dx\, e^{i\omega t} F(\omega)$$

can be closed in the complex ω-plane.

22. Ultrafast lasers have pulse widths that are on the order of 50 femptoseconds (a femptosecond is 10^{-15} seconds). This pulse width is difficult to measure since the signal is so short that it cannot be displayed directly on an oscilloscope. It is often measured using an autocorrelation technique with an apparatus similar to the one shown below:
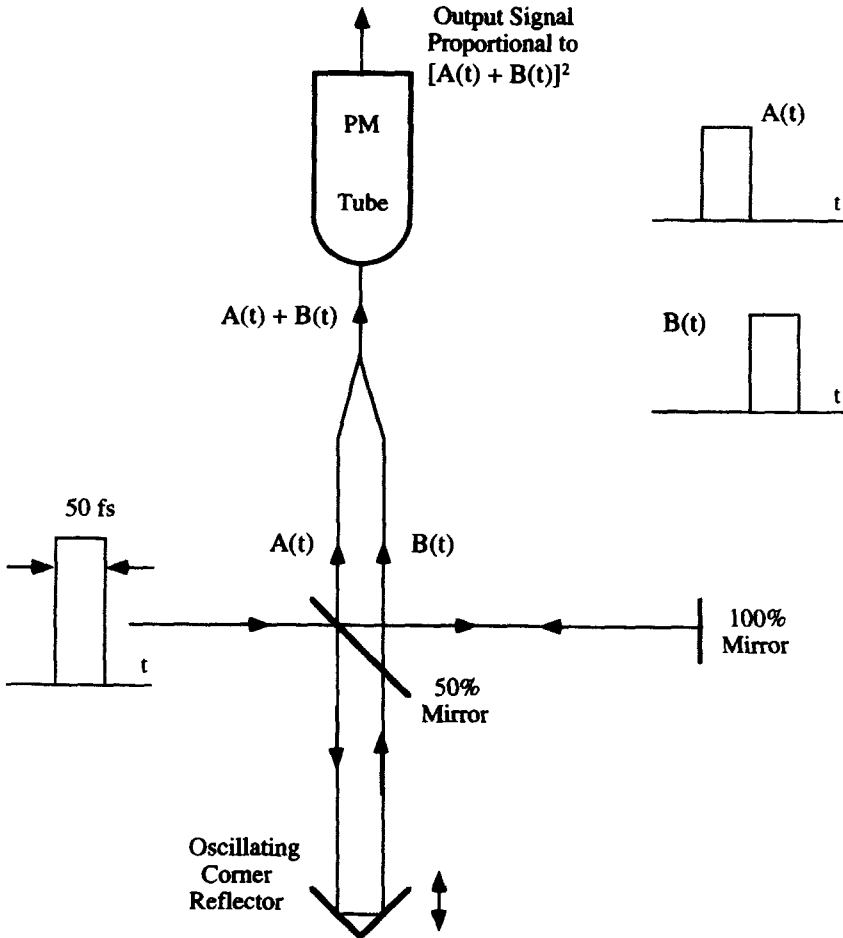
The original pulse is incident from the left on a 45°, 50% mirror. A 50% mirror reflects half of the incident signal and transmits the other half. The 100% mirror is at a fixed position and reflects all of the incident light. The corner reflector reflects all of the light back along the incident direction, regardless of the angle of incidence. This reflector is mounted on a stage that oscillates back and forth at a frequency of 10Hz. The split signal is then recombined at a detector, which is usually a photomultiplier tube or a fast LED detector. The detector generates an electrical signal which is proportional to the total number of photons in several pulses of the combined $[A(t) + B(t)]^2$ signal—i.e., it integrates the intensity of the signal over an interval long compared to the laser pulse width, but short compared to the period of the oscillating stage. Notice how the motion of the oscillating mirror causes the relative position of $B(t)$ compared to $A(t)$ to change. The output of the detector is applied to the vertical deflection of an oscilloscope,

while the signal that drives the position of the corner reflector is applied to the horizontal deflection.

(a) How far does light travel in 50 fs? For red light, how many oscillations of the electric field occur in 50 fs?

(b) Assume a train of 50 fs pulses is incident on this device, and one pulse is separated from the next by 0.01 μs. Assuming the motion of the oscillating reflector and the display of the oscilloscope are adjusted so the $B(t)$ signal moves from 200 fs before $A(t)$ to 200 fs after $A(t)$, what does the trace on the oscilloscope look like? To keep things simple, assume the pulse shape is square.

(c) How should the 100% mirror and the corner reflector be positioned with respect to the 50% mirror so that the operation in part (b) above is obtained?

(d) How is the width of the signal on the oscilloscope related to the pulse width of the incident signal? Treat the motion of the reflector as linear in time in the region where the two signals overlap.

9

LAPLACE TRANSFORMS

In this chapter, the Laplace transform is developed as a modification of the Fourier transform. The Laplace transform extends the realm of functions which can be transformed to include those which do not necessarily decay to zero as the independent variable grows to infinity. This is accomplished by introducing an exponentially decaying term in the transform operation. The inverse Laplace operation can be viewed as a modification of the inverse Fourier transform, with the Fourier contour moved off the real ω-axis into the complex $\underline{\omega}$-plane.

9.1 LIMITS OF THE FOURIER TRANSFORM

We can motivate the development of the Laplace transform by examining a function which clearly has no Fourier transform. Consider the ramp function shown in Figure 9.1 and described by

$$f(t) = \begin{cases} 0 & t < 0 \\ t & t > 0 \end{cases}.$$ (9.1)

This function does not go to zero as t goes to infinity and thus does not have a Fourier transform.

However, from the previous chapter, we know that a derivative operation in the time domain is equivalent to multiplication by the factor of $i\omega$ in the transformed frequency domain. In other words,

$$\mathcal{F}\left\{\frac{df(t)}{dt}\right\} = i\omega\underline{F}(\omega).$$ (9.2)
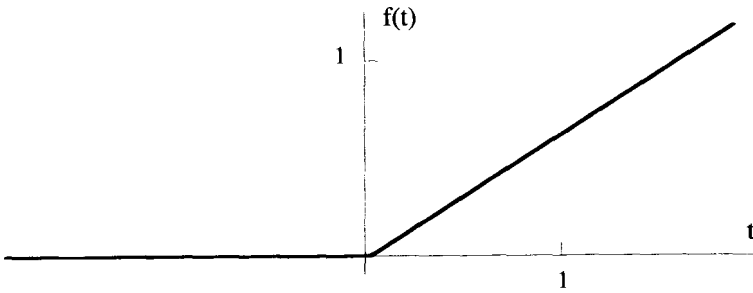
Figure 9.1   The Ramp Function

Notice that the ramp function given by Equation 9.1 is related, by two derivatives, to the delta function,

$$\frac{d^2 f(t)}{dt^2} = \delta(t), \tag{9.3}$$

as shown in Figure 9.2. In the previous chapter, we showed that the δ-function has the simple Fourier transform:

$$\mathcal{F}\{\delta(t)\} = \frac{1}{\sqrt{2\pi}}. \tag{9.4}$$
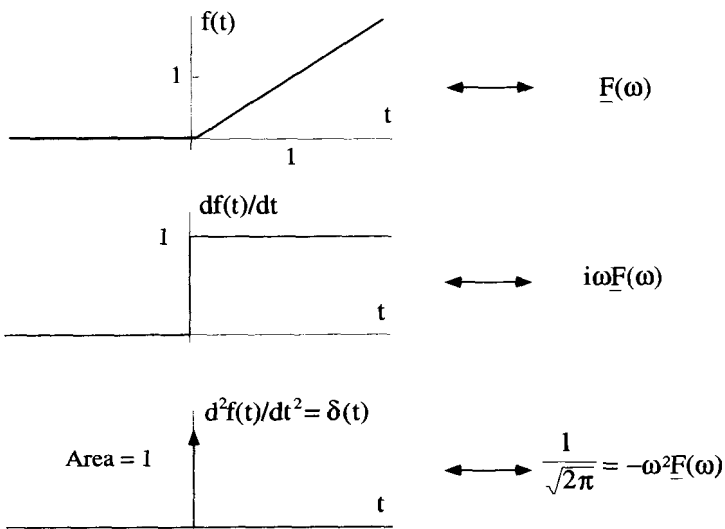
Figure 9.2   Relationship Between the Ramp and $\delta(t)$

So if we suspend our rigor for a moment, we can naively suppose Equation 9.2 lets us write the transform of the step function of Figure 9.2 as

$$\frac{-i}{\omega\sqrt{2\pi}}. \tag{9.5}$$

Applying Equation 9.2 one more time, with reckless abandon, gives the Fourier transform of the ramp function as

$$\frac{-1}{\omega^2\sqrt{2\pi}}. \tag{9.6}$$

Of course, we know neither the step function nor the ramp have legitimate Fourier transforms. This is confirmed by examining the inverse Fourier integral of the transform functions given in Equations 9.5 and 9.6. The inversion for Equation 9.5 is

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \, e^{i\omega t} \frac{-i}{\omega\sqrt{2\pi}}, \tag{9.7}$$

and the inversion for Equation 9.6 is

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \, e^{i\omega t} \frac{-1}{\omega^2\sqrt{2\pi}}. \tag{9.8}$$

Neither of these integrations can be performed because of the divergence of the integrands at $\omega = 0$. If you look at the representations of these integrals in the complex $\underline{\omega}$-plane, you will notice that a pole sits right on the Fourier contour.

This is the problem we examined briefly at the end of the previous chapter. By modifying the contour to dip below the offending pole, as shown in Figure 9.3, we actually can recover the original function from the transform function given in Equation 9.6. We will show this in the next section. Our sloppy derivation of the ramp function's transform must contain some grain of truth. By examining the inverse process a little more closely, we can arrive at the basis for the Laplace transform.
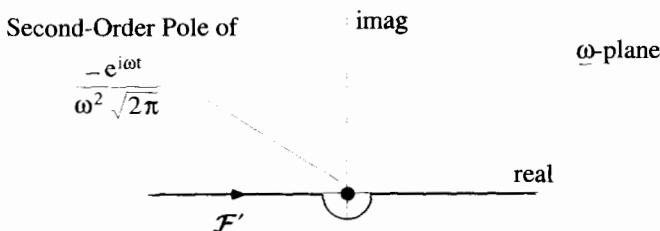
Figure 9.3   A Modified Fourier Contour for the Ramp Function

9.2   THE MODIFIED FOURIER TRANSFORM

In this section, we show how the standard Fourier transform can be modified by moving the inversion contour off the real axis. With this approach, we can transform functions that would not normally possess valid Fourier transforms. We start with some of the ideas introduced at the end of the previous chapter.

9.2.1   The Modified Transform Pair

In the previous section, we indicated that the ramp function of Equation 9.1 could be recovered from the transform given in Equation 9.6 using an inversion integral

$$ f(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathcal{F}'} d\underline{\omega} \; \frac{-e^{i\omega t}}{\underline{\omega}^2 \sqrt{2\pi}}, \tag{9.9} $$

on the modified contour $\mathcal{F}'$ shown in Figure 9.3. This contour lies along the real $\underline{\omega}$-axis, except at the origin where it dips below the pole. The discussion of the previous chapter shows that we can use closure in the upper half $\underline{\omega}$-plane for $t > 0$, and in the lower half for $t < 0$. Then, using the residue theorem, it is easy to show that

$$ f(t) \begin{cases} 0 & t < 0 \\ t & t > 0 \end{cases}. \tag{9.10} $$

The ramp is recovered by integrating the function given in Equation 9.6 along the modified Fourier contour $\mathcal{F}'$.

This, however, is not the only path that will recover the ramp function. Any path lying in the lower half $\underline{\omega}$-plane will work. Consider the path shown in Figure 9.4 that goes along the straight line $\omega_i = -\beta$, where ω_i is the imaginary part of $\underline{\omega}$:

$$ f(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathcal{F}''} d\underline{\omega} \; \frac{-e^{i\omega t}}{\underline{\omega}^2 \sqrt{2\pi}}. \tag{9.11} $$
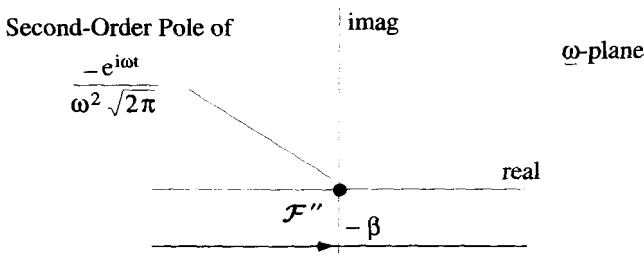
Figure 9.4   Another Modified Fourier Contour for the Ramp Function

Second-Order Pole

imag

ω'-plane

$\bullet\ \beta$

$\mathcal{F}$

real
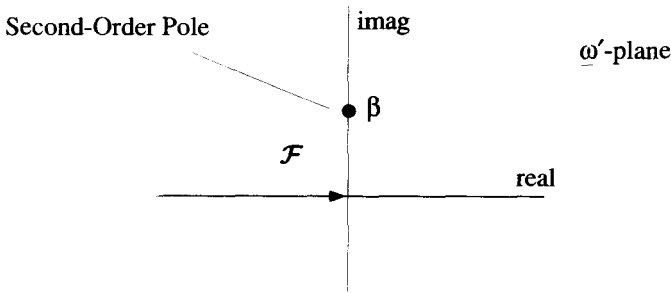
Figure 9.5 Inversion for the Ramp Function

This integration is easily performed by defining a transformation of variables with $\underline{\omega}' \equiv \underline{\omega} + i\beta$. With this transformation, the modified Fourier contour $\mathcal{F}''$ that was along $\omega_i = -\beta$ becomes a standard Fourier contour along the real ω_i' axis:

$$f(t) = \frac{e^{\beta t}}{2\pi} \int_{\mathcal{F}} d\underline{\omega}' \; \frac{-e^{i\underline{\omega}'t}}{(\underline{\omega}' - i\beta)^2}. \tag{9.12}$$

There is a second-order pole at $\underline{\omega}' = i\beta$, as shown in Figure 9.5. Now standard Fourier closure techniques can be applied to recover the same ramp function as before. Notice that, with the pole at $\underline{\omega}' = i\beta$, the $e^{\beta t}$ factor present in the numerator of Equation 9.12 is exactly canceled out by the residue. From these arguments it can be seen that the modified Fourier contour $\mathcal{F}'$ of Equation 9.9 can be placed anywhere in the lower half $\underline{\omega}$-plane, as shown in Figure 9.6.

The fact that the ramp function can be generated using Equation 9.6 in this inversion process raises the question as to whether this transform can be obtained directly from the ramp function itself, by extending the regular Fourier transform process into the complex $\underline{\omega}$-plane. If we allow ω to take on complex values, the Fourier transform becomes

$$\underline{F}(\underline{\omega}) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \; e^{-i\underline{\omega}t} f(t). \tag{9.13}$$

Second-order pole of
$\underline{F}(\underline{\omega})$

Imag

$\underline{\omega}$-plane

Real

Convergence region for
modified Fourier transform
of ramp function

$\mathcal{F}'$
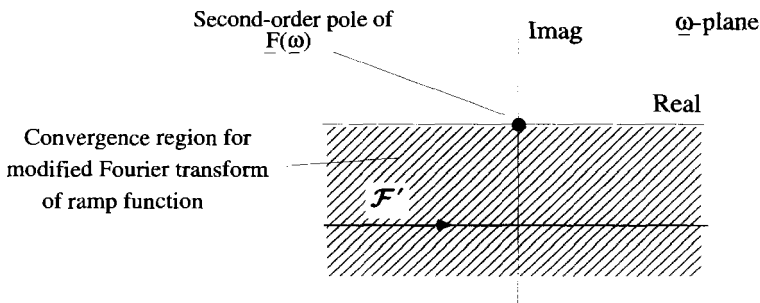
Figure 9.6 Convergence Region for the Ramp Transform

If $\underline{\omega} = \omega_r + i\omega_i$, we can rewrite Equation 9.13 as

$$\underline{F}(\underline{\omega}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \, e^{-i\omega_r t} \left[e^{\omega_i t} f(t) \right]. \tag{9.14}$$

Notice the integral in Equation 9.14 looks just like a regular Fourier transform of the function $f(t)e^{\omega_i t}$.

Can this process generate Equation 9.6 from the ramp function? Inserting the ramp function into Equation 9.14 gives

$$\underline{F}(\underline{\omega}) = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} dt \, e^{-i\underline{\omega} t} t. \tag{9.15}$$

The fact that the ramp function is zero for $t < 0$ makes the limits of integration range from 0 to $+\infty$. Integrating by parts, taking $u = t$ and $dv = e^{-i\underline{\omega} t} dt$, gives

$$\underline{F}(\underline{\omega}) = \frac{1}{\sqrt{2\pi}} \left[\frac{-te^{-i\underline{\omega} t}}{i\underline{\omega}} \right] \Bigg|_{0}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} dt \, \frac{e^{-i\underline{\omega} t}}{i\underline{\omega}}. \tag{9.16}$$

Both terms of this expression will converge only if $\omega_i < 0$. In that case, Equation 9.16 evaluates to

$$\underline{F}(\underline{\omega}) = \frac{-1}{\sqrt{2\pi}\underline{\omega}^2} \qquad \omega_i < 0. \tag{9.17}$$

Thus, at least for the ramp function, it appears we can define a modified Fourier transform pair by

$$\underline{F}(\underline{\omega}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt \, e^{-i\underline{\omega} t} f(t) \tag{9.18}$$

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathcal{F}'} d\underline{\omega} \, e^{i\underline{\omega} t} \underline{F}(\underline{\omega}), \tag{9.19}$$

with the requirement that the imaginary part of $\underline{\omega}$ be negative. The convergence region for this modified Fourier transform is identified in Figure 9.6.

9.2.2   Modified Fourier Transforms of a Growing Exponential

The modified Fourier transform pair given by Equations 9.18 and 9.19 has been shown to work for the ramp function. Will it work for other functions that do not possess regular Fourier transforms? Consider the exponentially growing function

$$f(t) = \begin{cases} 0 & t < 0 \\ e^{at} & t > 0 \end{cases}, \tag{9.20}$$
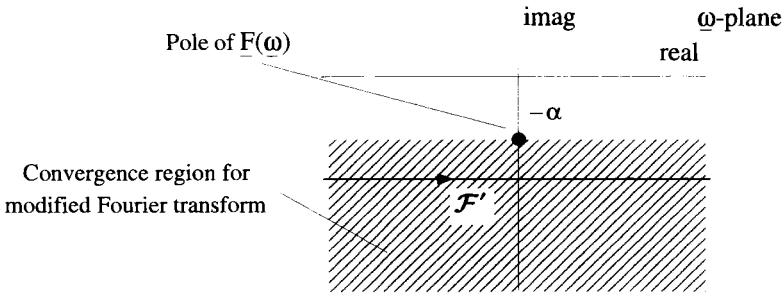
Figure 9.7   Convergence Region for Modified Fourier Transform of the Exponential Function

where α is a positive, real number. When this function is inserted into the modified transform of Equation 9.18, you can easily show, as long as $\omega_i < -\alpha$, that

$$\underline{F}(\omega) = \frac{-i}{\sqrt{2\pi}(\underline{\omega} + i\alpha)}. \tag{9.21}$$

Figure 9.7 shows this convergence region. Notice, as before, the convergence region is required to be below the pole of $\underline{F}(\underline{\omega})$.

Let's investigate the inverse operation, which must return the original function of Equation 9.20. Combining Equation 9.21 with the inversion integral of Equation 9.19 gives

$$f(t) = \frac{1}{2\pi i} \int_{\mathcal{F}'} d\underline{\omega} \; \frac{e^{i\underline{\omega} t}}{\underline{\omega} + i\alpha}. \tag{9.22}$$

The modified Fourier contour $\mathcal{F}'$ for this integration is shown in Figure 9.7.

The same stunt we used for the ramp function can be applied here. A simple variable transformation, $\underline{\omega}' = \underline{\omega} + i\beta$, converts $\mathcal{F}'$ to a contour on the real axis of the $\underline{\omega}'$-plane:

$$f(t) = \frac{e^{\beta t}}{2\pi i} \int_{\mathcal{F}} d\underline{\omega}' \; \frac{e^{i\underline{\omega}' t}}{\underline{\omega}' + i\alpha - i\beta} \tag{9.23}$$

Again, the integral in Equation 9.23 can be evaluated by closing the contour. The complex exponential forces us to close in the lower half plane for $t < 0$ and in the upper half plane for $t > 0$. There is a single pole at $\omega' = i(\beta - \alpha)$ with residue $e^{\alpha t} e^{-\beta t}$. Thus, the residue theorem tells us

$$f(t) = \begin{cases} 0 & t < 0 \\ e^{\alpha t} & t > 0 \end{cases}, \tag{9.24}$$

which is indeed the original function.

9.2.3   Closure for Contours off the Real Axis

In the previous two examples, we made a variable change from ω to ω', so the $\mathcal{F}'$ contour was changed into a contour along the real axis. The only reason we did this was so we could use the standard arguments developed in Chapters 6 and 8 to show the contribution of the closing contour was zero. Actually, this step is not necessary. We can close directly in the ω-plane, if we analyze the contribution from the closing contour a little more carefully. This is more work than the variable transformation approach, but once this general argument is understood, it simplifies the whole process.

To see how this works, let's return to the specific integration of Equation 9.22. For $t < 0$, consider the closed contour shown in Figure 9.8. In the limit of $R_o \to \infty$, the straight line segment from A to B is the modified Fourier contour $\mathcal{F}'$ of Equation 9.22. This contour can be closed, as shown, with the circular segment in the lower ω-plane. Using the standard closure arguments, it is easy to see the contribution from this circular segment as $R_o \to \infty$ is zero. Because there are no poles of the integrand of Equation 9.22 below $\mathcal{F}'$ we obtain

$$f(t) = 0 \qquad t < 0. \tag{9.25}$$

For $t > 0$ consider the closed, solid contour shown in Figure 9.9. In the limit of $R_o \to \infty$, the straight line-segment AB again is $\mathcal{F}'$, the modified Fourier contour. The contour is now closed by a circular segment $BCDA$ that is mostly in the upper half ω-plane, but has sections, BC and DA, that extend into the lower half ω-plane. The contribution from this $BCDA$ circular segment is also zero, but because of the sections BC and DA this fact is more difficult to show.
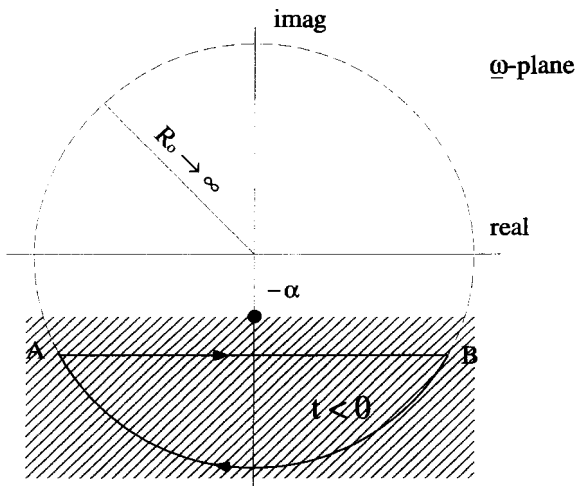
Figure 9.8   Closure for the Modified Fourier Transform for $t < 0$
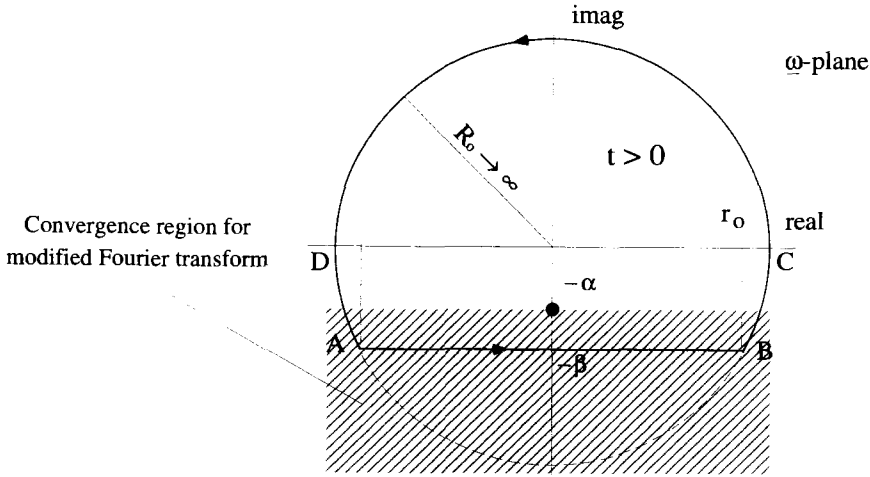
Figure 9.9   Closure for the Modified Fourier Transform for $t > 0$

To show that this contribution is zero, we begin by identifying the points A and B. The point A is located at $\underline{\omega} = -r_o - i\beta$, and B is located at $\underline{\omega} = r_o - i\beta$, as shown in Figure 9.9. Notice that as $R_o \to \infty$, $r_o \to R_o$, while β remains fixed. The integration along the total circular segment can be integrated by parts:

$$\frac{1}{2\pi i} \int_{BCDA} d\underline{\omega}\ \frac{e^{i\underline{\omega} t}}{\underline{\omega} + i\alpha} = \left. \frac{-e^{i\underline{\omega} t}}{2\pi t(\underline{\omega} + i\alpha)} \right|_B^A - \int_{BCDA} d\underline{\omega}\ \frac{e^{i\underline{\omega} t}}{2\pi t(\underline{\omega} + i\alpha)^2}. \quad (9.26)$$

Evaluating the first term on the RHS gives

$$\left. \frac{-e^{i\underline{\omega} t}}{2\pi t(\underline{\omega} + i\alpha)} \right|_B^A = \frac{e^{\beta t}}{2\pi t} \left[\frac{e^{-itr_o}}{-r_o + i(\alpha - \beta)} - \frac{e^{itr_o}}{r_o + i(\alpha - \beta)} \right]. \quad (9.27)$$

As R_o and $r_o \to \infty$, the RHS of Equation 9.27 goes to zero. The integration along the circular segment can therefore be broken up into three parts:

$$\frac{1}{2\pi i} \int_{BCDA} d\underline{\omega}\ \frac{e^{i\underline{\omega} t}}{\underline{\omega} + i\alpha} = -\int_{BC} d\underline{\omega}\ \frac{e^{i\underline{\omega} t}}{2\pi t(\underline{\omega} + i\alpha)^2}$$

$$-\int_{CD} d\underline{\omega}\ \frac{e^{i\underline{\omega} t}}{2\pi t(\underline{\omega} + i\alpha)^2} \quad (9.28)$$

$$-\int_{DA} d\underline{\omega}\ \frac{e^{i\underline{\omega} t}}{2\pi t(\underline{\omega} + i\alpha)^2}.$$

The second integral, on segment CD, goes to zero as $R_o \to \infty$ by the standard arguments used for the Fourier transform. This leaves only the integrals along segments
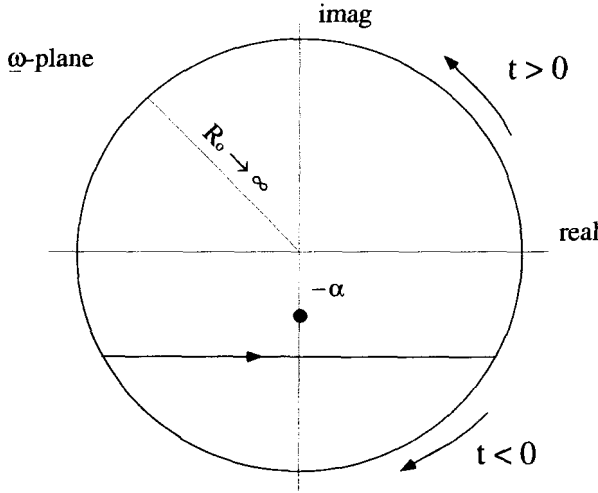
Figure 9.10   Closure of a Modified Fourier Contour

BC and DA. Consider the BC segment. If $\underline{\omega} = \omega_r + i\omega_i$, everywhere along this segment $\omega_r \geq r_o$ and $\omega_i \geq -\beta$. This implies

$$\left| \frac{e^{i\underline{\omega}t}}{(\underline{\omega} + i\alpha)^2} \right| = \left| \frac{e^{it(\omega_r + i\omega_i)}}{[\omega_r + i(\alpha + \omega_i)]^2} \right| \leq \frac{e^{\beta t}}{R_o^2}. \tag{9.29}$$

As $R_o \rightarrow \infty$, the integrand is dropping off as $1/R_o^2$ while the path length is limited to a length of β. Consequently, as $R_o \rightarrow \infty$, the contribution from the BC section, for any fixed value of t, goes to zero. The same arguments can be made for the DA section. Therefore, for $t > 0$, the integration along the circular segment BCDA goes to zero and we can close the contour. The pole at $\underline{\omega} = -i\alpha$ is inside the contour of Figure 9.9, and the residue theorem gives

$$f(t) = e^{\alpha t} \qquad t > 0. \tag{9.30}$$

To summarize, we can close the modified Fourier contour for the growing exponential function with the closed contours shown in Figure 9.10. This approach gives the same result as was obtained by first transforming to the $\underline{\omega}'$-plane. In the rest of this chapter, we will take this kind of reasoning for granted as we close different contours.

9.2.4   General Properties of the Modified Fourier Transform

The ramp and exponential function examples demonstrate the general conditions which allow the modified Fourier transform to exist. First, the function to be transformed must be exponentially bounded. The meaning of that expression is evident

from looking at Equation 9.14. It must be possible to find an ω_i such that $e^{\omega_i t} f(t)$ goes to zero as t goes to infinity. If this is the case, the function $e^{\omega_i t} f(t)$ will have a regular Fourier transform. Second, the function we want to transform must be zero for $t < 0$. Actually, it is quite easy to develop the modified Fourier transform arguments so that $f(t)$ must be zero for all $t < t_o$, where t_o is any real number. You can explore this idea further in one of the exercises at the end of this chapter.

The modified Fourier transform of Equation 9.18 will always exist for values of $\underline{\omega}$ with imaginary parts less than the imaginary parts of *all* the poles of $\underline{F}(\underline{\omega})$. This implies the $\mathcal{F}'$ inversion contour of Equation 9.19 must also lie below all these poles. The inversion can be accomplished by closing above the modified Fourier contour for $t > 0$ and below it for $t < 0$.

9.3   THE LAPLACE TRANSFORM

The standard Laplace transform is really just the modified Fourier transform we have just discussed with a variable change and a rearrangement of the constants. The variable change involves the simple complex substitution

$$\underline{s} = i\underline{\omega}. \tag{9.31}$$

Because the inversion equation is frequently evaluated using the residue theorem, the constants are juggled around so the leading constant of the inversion is $1/(2\pi i)$. With these changes, the Laplace transform pair follows directly from Equations 9.18 and 9.19:

$$\underline{F}(\underline{s}) = \int_0^\infty dt\, e^{-\underline{s}t} f(t) \tag{9.32}$$

$$f(t) = \frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s}\, e^{\underline{s}t} \underline{F}(\underline{s}). \tag{9.33}$$

It is implicitly assumed by the form of Equation 9.32 that $f(t) = 0$ for all $t < 0$. These equations are sometimes written with a shorthand operator notation as:

$$\underline{F}(\underline{s}) = \mathcal{L}\{f(t)\} \tag{9.34}$$

$$f(t) = \mathcal{L}^{-1}\{\underline{F}(\underline{s})\}. \tag{9.35}$$

The picture of the Laplace inversion in the complex $\underline{s}$-plane looks very much like the pictures of the modified Fourier transform, except for a rotation of $\pi/2$. Figure 9.11 shows a typical convergence region and inversion contours. The Laplace transform converges in the region to the right of all the poles of $\underline{F}(\underline{s})$. The Laplace inversion contour $\mathcal{L}$, which is called the Laplace or Bromwich contour, must lie in this convergence region and extends from $\underline{s} = \beta - i\infty$ to $\underline{s} = \beta + i\infty$. The inversion contour is almost always evaluated by closing the complex integral in Equation 9.33. Using similar arguments given for the modified Fourier transform, the Laplace contour is closed to the right for $t < 0$ and to the left for $t > 0$. Because $\mathcal{L}$
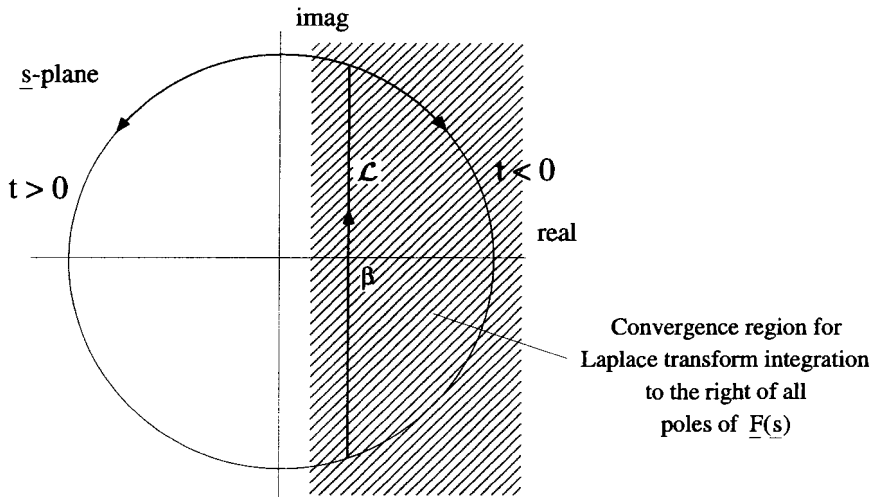
Figure 9.11   Complex $\underline{s}$-Plane for the Laplace Inversion

lies to the right of all the poles, the inversion always generates an $f(t)$ that is zero for $t < 0$.

It is the real part of $\underline{s}$ that allows the Laplace transform to converge for functions which do not have valid Fourier transforms. For an exponentially growing $f(t)$, as long as the real part of $\underline{s}$ is larger than the exponential growth rate of $f(t)$, the Laplace transform will converge. If $f(t)$ is not exponentially growing, but in fact exponentially decaying, all the poles of $\underline{F}(\underline{s})$ are to the left of the imaginary axis, and the convergence region will include the imaginary $\underline{s}$-axis. In a case like this, the entire Laplace contour can be placed on the imaginary $\underline{s}$-axis, and the Laplace transform is equivalent to a regular Fourier transform, with the exception of the fact that $f(t) = 0$ for $t < 0$. This particular limitation will be removed when we consider the bilateral Laplace transform at the end of this chapter. Thus, we can really view the Fourier transform as a special case of the more general Laplace transform.

9.4   LAPLACE TRANSFORM EXAMPLES

In this section, the Laplace transform of several functions will be determined. With the exception of the δ-function example, these functions do not have Fourier transforms.

9.4.1   The δ-Function

The function $\delta(t - t_o)$ has a Fourier transform, and if $t_o > 0$, it is zero for $t < 0$. Therefore, we should expect its Laplace transform, except for an overall constant, to be the same as its Fourier transform with $-i\underline{s}$ substituted for ω. This is easily

confirmed by performing the transform:

$$\mathcal{L}\{\delta(t - t_o)\} = \int_0^\infty dt\ e^{-st}\delta(t - t_o)$$

$$= e^{-st_o}. \tag{9.36}$$

There is no limit on the real part of $\underline{s}$ for performing this integration, and therefore the region of convergence covers the entire complex $\underline{s}$-plane. Consequently, for the inversion of the Laplace transform of the δ-function, the Laplace contour can be placed anywhere. That is, any value of β is acceptable. Incidentally, the fact that we can let the Laplace contour lie on the imaginary $\underline{s}$-axis is the reason the δ-function possesses a Fourier transform.

The inverse transform that returns the δ-function is

$$\mathcal{L}^{-1}\{e^{-st_o}\} = \frac{1}{2\pi i}\int_{\mathcal{L}} d\underline{s}\ e^{st}e^{-st_o}, \tag{9.37}$$

with the Laplace contour $\mathcal{L}$ located as shown in Figure 9.12.

This is one of the few cases where the Laplace inversion cannot be done by closing the contour. Notice, however, because the convergence region is the entire complex $\underline{s}$-plane, the Laplace contour can be placed on the imaginary $\underline{s}$-axis. In that case, $\underline{s} = i\omega$ and $d\underline{s} = id\omega$, so the RHS of Equation 9.37 becomes

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\ e^{i\omega(t-t_o)} = \delta(t - t_o). \tag{9.38}$$
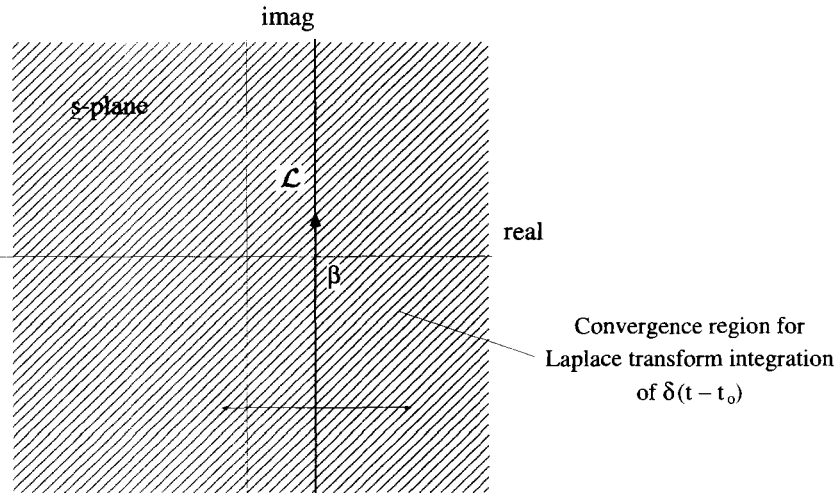
Figure 9.12   Laplace Inversion of the δ-Function
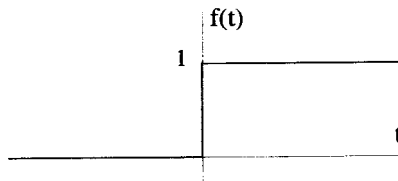
Figure 9.13 The Heaviside Step Function

This final result comes from the orthogonality relation derived in the previous chapter. Equation 9.37 is actually a more general form of that orthogonality relation:

$$\frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s}\, e^{\underline{s}(t-t_o)} = \delta(t - t_o). \tag{9.39}$$

9.4.2 The Heaviside Function

The Heaviside step function does not possess a Fourier transform. It does, however, have a Laplace transform. Recall the definition of the Heaviside function is

$$f(t) = \begin{cases} 0 & t < 0 \\ 1 & t > 0 \end{cases}, \tag{9.40}$$

which is shown in Figure 9.13.

The Laplace transform of this function is

$$\int_0^\infty dt\, e^{-st} = \frac{1}{\underline{s}} \qquad s_r > 0. \tag{9.41}$$
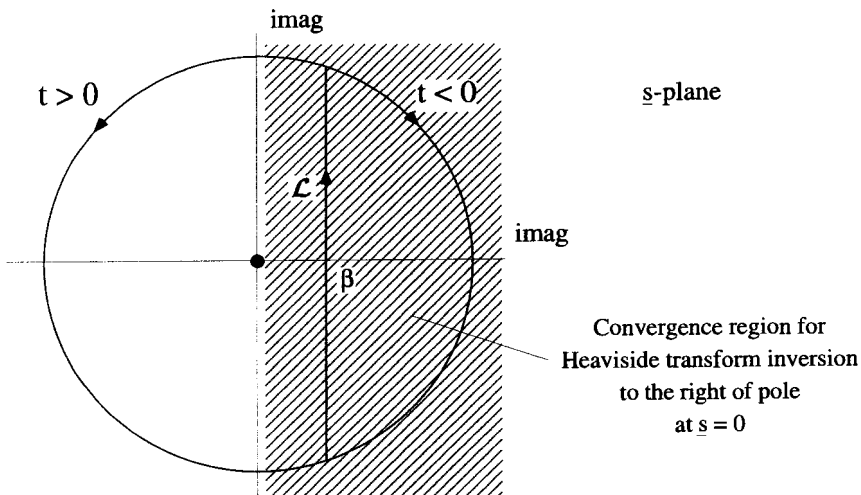
Figure 9.14 Inversion of the Heaviside Laplace Transform

This integration can be performed only when the real part of $\underline{s}$, which we call s_r, is positive. The Laplace inversion of this transform is

$$\frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s}\, e^{\underline{s}t} \frac{1}{\underline{s}}. \tag{9.42}$$

The integrand has a first-order pole at $\underline{s} = 0$, with a residue of 1. To be within the convergence region, the Laplace contour must fall to the right of this pole. It can be closed on the left for $t > 0$ and to the right for $t < 0$, as shown in Figure 9.14. The evaluation in either case is straightforward and recovers the original step function given by Equation 9.40.

This example shows why the operation described in Figure 8.43 of the previous chapter worked. By deforming the Fourier contour to $\mathcal{F}'$, as shown in that figure, the operation was really a modified Fourier transform, equivalent to the Laplace transform of this example.

9.4.3 An Exponentially Growing Sinusoid

Consider the exponentially growing sinusoid

$$f(t) = \begin{cases} 0 & t < 0 \\ e^{\alpha_o t}\sin(\omega_o t) & t > 0 \end{cases}, \tag{9.43}$$

with $\alpha_o > 0$, as shown in Figure 9.15. Again, this function does not possess a traditional Fourier transform. Its Laplace transform, however, is given by

$$\underline{F}(\underline{s}) = \int_0^\infty dt\, e^{-\underline{s}t} e^{\alpha_o t}\sin(\omega_o t). \tag{9.44}$$

This integral exists only when $s_r > \alpha_o$. In that region,

$$\underline{F}(\underline{s}) = \frac{\omega_o}{(\underline{s} - \underline{s}_1)(\underline{s} - \underline{s}_2)} \qquad s_r > \alpha_o, \tag{9.45}$$
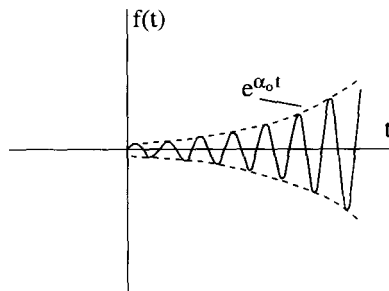
Figure 9.15 The Exponentially Growing Sinusoid
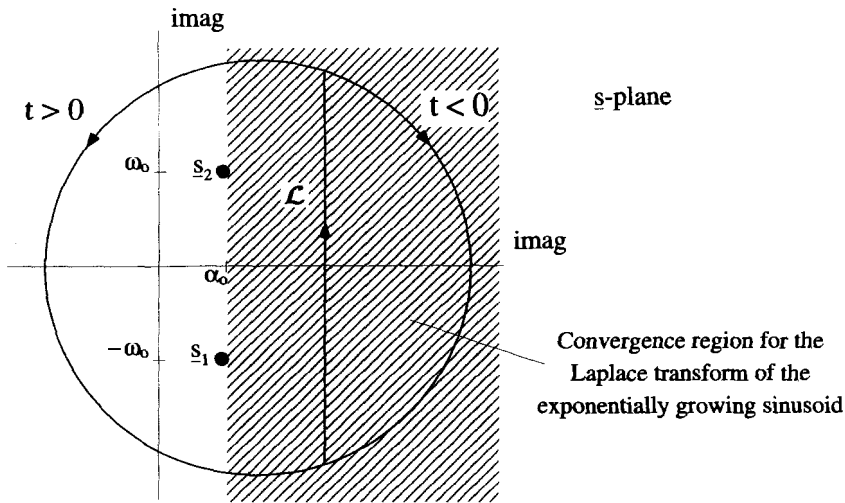
Figure 9.16   Inversion of the Laplace Transform for the Exponentially Growing Sinusoid

where $\underline{s}_1 \equiv \alpha_o - i\omega_o$ and $\underline{s}_2 \equiv \alpha_o + i\omega_o$. The inversion contour must fall to the right of the two poles at $\underline{s}_1$ and $\underline{s}_2$, as shown in Figure 9.16. The integration for the inversion,

$$\frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s}\, e^{\underline{s}t} \frac{\omega_o}{(\underline{s} - \underline{s}_1)(\underline{s} - \underline{s}_2)} = \begin{cases} 0 & t < 0 \\ e^{\alpha_o t} \sin(\omega_o t) & t > 0 \end{cases}, \qquad (9.46)$$

generates the original exponentially growing sinusoid, as you will prove in an exercise at the end of this chapter.

This example should be compared with the operation described in Figure 8.47 of the previous chapter. We can now see why modifying the Fourier contour to dip below the poles gave the correct answer. Unwittingly, we were performing a Laplace transform!

9.5   PROPERTIES OF THE LAPLACE TRANSFORM

As we did with the properties of the Fourier transform, we will discuss the properties of the Laplace transform using the two functions $f(t)$ and $g(t)$. We assume both functions are zero for $t < 0$, and that they possess the Laplace transforms

$$\underline{F}(\underline{s}) = \int_0^\infty dt\, e^{-\underline{s}t} f(t) \qquad (9.47)$$

and

$$\underline{G}(\underline{s}) = \int_0^\infty dt\, e^{-\underline{s}t} g(t). \qquad (9.48)$$

These transform exist for the real part of $\underline{s}$ greater than some value, not necessarily the same for each function. Both these transforms can be inverted by integration along some Laplace contour

$$f(t) = \frac{1}{2\pi i} \int_{\mathcal{L}_f} d\underline{s} \, e^{\underline{s}t} \underline{F}(\underline{s}) \tag{9.49}$$

and

$$g(t) = \frac{1}{2\pi i} \int_{\mathcal{L}_g} d\underline{s} \, e^{\underline{s}t} \underline{G}(\underline{s}). \tag{9.50}$$

We use two different contours, $\mathcal{L}_f$ and $\mathcal{L}_g$, to allow for the possibility of two different convergence regions.

9.5.1   Delay

Consider the function $f(t - t_o)$, which is the function $f(t)$ delayed by t_o. The Laplace transform of this delayed function is

$$\mathcal{L}\{f(t - t_o)\} = \int_0^\infty dt \, e^{-\underline{s}t} f(t - t_o). \tag{9.51}$$

Make the substitution $\alpha = t - t_o$, so Equation 9.51 becomes

$$\begin{aligned}
\mathcal{L}\{f(t - t_o)\} &= \int_{-t_o}^\infty d\alpha \, e^{-\underline{s}(\alpha + t_o)} f(\alpha) \\
&= e^{-\underline{s}t_o} \int_0^\infty d\alpha \, e^{-\underline{s}\alpha} f(\alpha) \tag{9.52} \\
&= e^{-\underline{s}t_o} \underline{F}(\underline{s}).
\end{aligned}$$

The change in the lower limit in the second step is valid because $f(\alpha) = 0$ for $\alpha < 0$. A delay in time is equivalent to multiplying the Laplace transform by an exponential factor.

9.5.2   Differentiation

The Laplace transform of $df(t)/dt$ is

$$\mathcal{L}\left\{ \frac{df(t)}{dt} \right\} = \int_0^\infty dt \, e^{-\underline{s}t} \frac{df(t)}{dt}. \tag{9.53}$$

The RHS of this expression can be integrated by parts with $\underline{u} = e^{-\underline{s}t}$ and $dv = (df(t)/dt) \, dt$ to give

$$\mathcal{L}\left\{ \frac{df(t)}{dt} \right\} = f(t)e^{-\underline{s}t} \Big|_0^\infty + \int_0^\infty dt \, e^{-\underline{s}t} \underline{s} f(t). \tag{9.54}$$

If the real part of $\underline{s}$ is large enough to make $f(t)e^{-\underline{s}t}$ go to zero as $t \rightarrow \infty$, then

$$\mathcal{L}\left\{\frac{df(t)}{dt}\right\} = -f(0) + \underline{s}\,\underline{F}(\underline{s}). \qquad (9.55)$$

Notice the difference between this result and the derivative relationship for the Fourier transform.

9.5.3   Convolution

Now consider a function $h(t)$, which is the convolution of $f(t)$ with $g(t)$:

$$h(t) = \int_{-\infty}^{\infty} d\tau\, f(\tau)g(t - \tau). \qquad (9.56)$$

The Laplace transform of $h(t)$ can be written as

$$\begin{aligned}
\underline{H}(\underline{s}) &= \int_{0}^{\infty} dt\, e^{-\underline{s}t} h(t) \\
&= \int_{-\infty}^{\infty} dt\, e^{-\underline{s}t} h(t) \\
&= \int_{-\infty}^{\infty} dt\, e^{-\underline{s}t} \int_{-\infty}^{\infty} d\tau\, f(\tau)g(t - \tau). \qquad (9.57)
\end{aligned}$$

In the second step, we extended the lower limit to $-\infty$ using the result of Exercise 8.15, which says if $f(t) = 0$ and $g(t) = 0$ for $t < 0$, then $h(t) = 0$ for $t < 0$. Now substitute the inverse Laplace transforms for $f(\tau)$ and $g(t - \tau)$ to convert the RHS of Equation 9.57 into

$$\int_{-\infty}^{\infty} dt\, e^{-\underline{s}t} \int_{-\infty}^{\infty} d\tau\, \underbrace{\left[\frac{1}{2\pi i}\int_{\mathcal{L}_f} d\underline{s}'\, e^{\underline{s}'\tau}\underline{F}(\underline{s}')\right]}_{f(\tau)} \underbrace{\left[\frac{1}{2\pi i}\int_{\mathcal{L}_g} d\underline{s}''\, e^{\underline{s}''(t-\tau)}\underline{G}(\underline{s}'')\right]}_{g(t-\tau)}. \quad (9.58)$$

We have been careful to keep the inversion integrations independent by using two distinct variables of integration, $\underline{s}'$ and $\underline{s}''$. Notice, the inversion for $f(\tau)$ has been done on a Laplace contour $\mathcal{L}_f$, which lies within the convergence region for the transform of $f(t)$. The integration variable $\underline{s}'$ lies on $\mathcal{L}_f$. Similarly, the inversion for $g(t - \tau)$ has been done on a contour $\mathcal{L}_g$, which must lie within the convergence region for $\underline{G}(\underline{s})$. The integration variable $\underline{s}''$ lies on $\mathcal{L}_g$. Figure 9.17 shows this situation, with the $\mathcal{L}_f$ and $\mathcal{L}_g$ contours placed as far to the left in their respective convergence regions as possible. Remember, the contours are not necessarily in the same location because the convergence regions of the two transforms may be different. The convergence region for $\underline{H}(\underline{s})$ falls to the right of all the poles of $\underline{H}(\underline{s})$. We will show that the poles of $\underline{H}(\underline{s})$ are just a combination of the poles of $\underline{F}(\underline{s})$ and $\underline{G}(\underline{s})$, so the convergence region
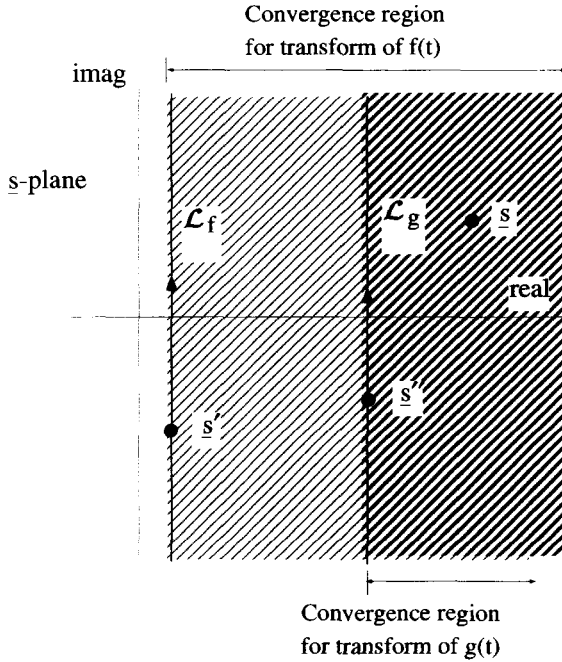
Figure 9.17   Convergence Regions and Laplace Contours for $f(t)$ and $g(t)$

of $\underline{H}(\underline{s})$ is just the overlap of the convergence regions of $\underline{F}(\underline{s})$ and $\underline{G}(\underline{s})$. The variable $\underline{s}$ of Equation 9.58 lies anywhere in this convergence region of $\underline{H}(\underline{s})$.

Because $\mathcal{L}_f$ is to the right of all the poles of $\underline{F}(\underline{s})$, and $\mathcal{L}_g$ is to the right of all the poles of $\underline{G}(\underline{s})$, contour deformation allows us to move both $\mathcal{L}_f$ and $\mathcal{L}_g$ to the right without changing the values of their integrals. It will be very convenient if we move them to the right so that they become a common contour $\mathcal{L}$ that passes through the point $\underline{s}$. Remember $\underline{s}$ is the point where we want to evaluate $\underline{H}(\underline{s})$. Figure 9.18 shows this new configuration. In this situation, $\underline{s}$, $\underline{s}'$, and $\underline{s}''$ will all have the same real part which we will call s_{ro}. For the integrations in Equation 9.58, s_{ro} is fixed, but the imaginary parts of $\underline{s}'$ and $\underline{s}''$, which we call s_i' and s_i'', vary independently.

Using the common Laplace contour and moving the $\int dt$ operator to the right in Equation 9.58 gives

$$\frac{-1}{4\pi^2} \int_{-\infty}^{\infty} d\tau \left[\int_{\mathcal{L}} d\underline{s}' e^{\underline{s}'\tau} \underline{F}(\underline{s}') \right] \left[\int_{\mathcal{L}} d\underline{s}'' e^{-\underline{s}''\tau} \underline{G}(\underline{s}'') \int_{-\infty}^{\infty} dt\, e^{(\underline{s}''-\underline{s})t} \right]. \qquad (9.59)$$

Notice what happens to the last integral in this equation. Because $\underline{s}''$ and $\underline{s}$ have the same real part, that is $\underline{s} = s_{ro} + is_i$ and $\underline{s}'' = s_{ro} + is_i''$, this integral is just the orthogonality relation:

$$\int_{-\infty}^{\infty} dt\, e^{it(s_i''-s_i)} = 2\pi\delta(s_i'' - s_i). \qquad (9.60)$$
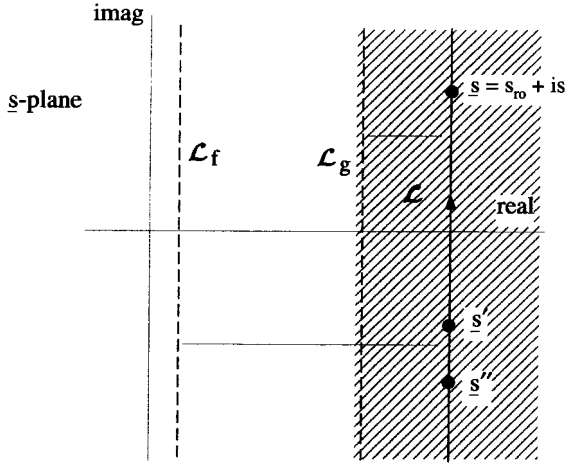
Figure 9.18   The Common Region of Convergence and Laplace Contour

Applying this to expression 9.59 gives

$$\frac{-i}{2\pi} \int_{-\infty}^{\infty} d\tau \left[\int_{\mathcal{L}} d\underline{s}' e^{\underline{s}'\tau} \underline{F}(\underline{s}') \right] \left[\int_{-\infty}^{\infty} ds_i'' e^{-\underline{s}''\tau} \underline{G}(\underline{s}'')\delta(s_i'' - s_i) \right], \qquad (9.61)$$

where we have also made use of the fact that along $\mathcal{L}$, $d\underline{s}'' = ids_i''$, with s_i'' ranging from minus to plus infinity. The δ-function makes the integral over $\underline{s}''$ very easy and lets us write

$$\underline{H}(\underline{s}) = \frac{-i}{2\pi} \int_{-\infty}^{\infty} d\tau \left[e^{-\underline{s}\tau} \underline{G}(\underline{s}) \int_{\mathcal{L}} d\underline{s}' e^{\underline{s}'\tau} \underline{F}(\underline{s}') \right]. \qquad (9.62)$$

Interchange the remaining two integrals:

$$\underline{H}(\underline{s}) = \frac{-i}{2\pi} \int_{\mathcal{L}} d\underline{s}' \left[\underline{F}(\underline{s}')\underline{G}(\underline{s}) \int_{-\infty}^{\infty} d\tau \, e^{\tau(\underline{s}'-\underline{s})} \right]. \qquad (9.63)$$

We can now pull a similar stunt to the one used on the last integral in Equation 9.59. Because $\underline{s}'$ and $\underline{s}$ have the same real part, this looks like another instance of the orthogonality relation:

$$\int_{-\infty}^{\infty} d\tau \, e^{i\tau(s_i'-s_i)} = 2\pi\delta(s_i' - s_i). \qquad (9.64)$$

Again making use of the fact that $d\underline{s}' = ids_i'$ along the $\mathcal{L}$ contour, we can write

$$\underline{H}(\underline{s}) = \int_{-\infty}^{\infty} ds_i' \left[\underline{F}(\underline{s}')\underline{G}(\underline{s})\delta(s_i' - s_i) \right] \qquad (9.65)$$

$$= \underline{F}(\underline{s})\underline{G}(\underline{s}). \qquad (9.66)$$

This is an elegant final result. The Laplace transform of the convolution of two functions is the product of the Laplace transforms of the functions

$$\mathcal{L}\{f(t) \otimes g(t)\} = \underline{F}(\underline{s})\underline{G}(\underline{s}). \tag{9.67}$$

Because $\underline{H}(\underline{s})$ is the product of $\underline{F}(\underline{s})$ and $\underline{G}(\underline{s})$, the poles of $\underline{H}(\underline{s})$ are just a combination of the poles of $\underline{F}(\underline{s})$ and $\underline{G}(\underline{s})$. The convergence region for $\underline{H}(\underline{s})$ is therefore the intersection of the convergence regions of both $\underline{F}(\underline{s})$ and $\underline{G}(\underline{s})$, as we assumed earlier.

9.5.4 Multiplication

As a final situation, consider the product of two functions

$$h(t) = f(t)g(t). \tag{9.68}$$

The Laplace transform of $h(t)$ is given by

$$\underline{H}(\underline{s}) = \int_0^\infty dt\, e^{-\underline{s}t} f(t)g(t) \tag{9.69}$$

$$= \int_{-\infty}^\infty dt\, e^{-\underline{s}t} \left[\frac{1}{2\pi i} \int_{\mathcal{L}_f} d\underline{s}' e^{\underline{s}'t} \underline{F}(\underline{s}') \right] \left[\frac{1}{2\pi i} \int_{\mathcal{L}_g} d\underline{s}'' e^{\underline{s}''t} \underline{G}(\underline{s}'') \right].$$

Again $f(t)$ and $g(t)$ have been assumed to be zero for $t < 0$, so that the dt integration of the Laplace transform can be extended to negative infinity. As with the previous case, the Laplace contours $\mathcal{L}_f$ and $\mathcal{L}_g$ are in their respective convergence regions and, initially, are placed as far to the left as possible, as shown in Figures 9.17 and 9.19.

We begin the manipulation of Equation 9.69 by interchanging the order of integration, to obtain:

$$\underline{H}(\underline{s}) = -\frac{1}{4\pi^2} \int_{\mathcal{L}_f} d\underline{s}'\, \underline{F}(\underline{s}') \int_{\mathcal{L}_g} d\underline{s}''\, \underline{G}(\underline{s}'') \int_{-\infty}^\infty dt\, e^{t(\underline{s}'+\underline{s}''-\underline{s})}. \tag{9.70}$$

On $\mathcal{L}_f$ let $\underline{s}' = s'_{ro} + is'_i$ and on $\mathcal{L}_g$ let $\underline{s}'' = s''_{ro} + is''_i$ where s'_{ro} and s''_{ro} are constants. If $\underline{s} = s_r + is_i$, then in order for the dt integral to exist, we must have

$$s'_{ro} + s''_{ro} - s_r = 0. \tag{9.71}$$

If the condition is not met, the dt integral diverges and $\underline{H}(\underline{s})$ does not exist. If this condition is met, however, the dt integral becomes a δ-function,

$$\int_{-\infty}^\infty dt\, e^{t(\underline{s}'+\underline{s}''-\underline{s})} = \int_{-\infty}^\infty dt\, e^{it(s'_i+s''_i-s_i)} \tag{9.72}$$

$$= 2\pi\delta(s'_i + s''_i - s_i), \tag{9.73}$$

and $\underline{H}(\underline{s})$ can be evaluated.

The condition of Equation 9.71 constrains the convergence region of $\underline{H}(\underline{s})$. If $\underline{F}(s')$ converges for $s'_r > s_f$ and $\underline{G}(\underline{s}'')$ converges for $s''_r > s_g$, then Equation 9.71 says $H(\underline{s})$ can exist only for $s_r > s_f + s_g$. This situation is shown in Figure 9.19. To see why
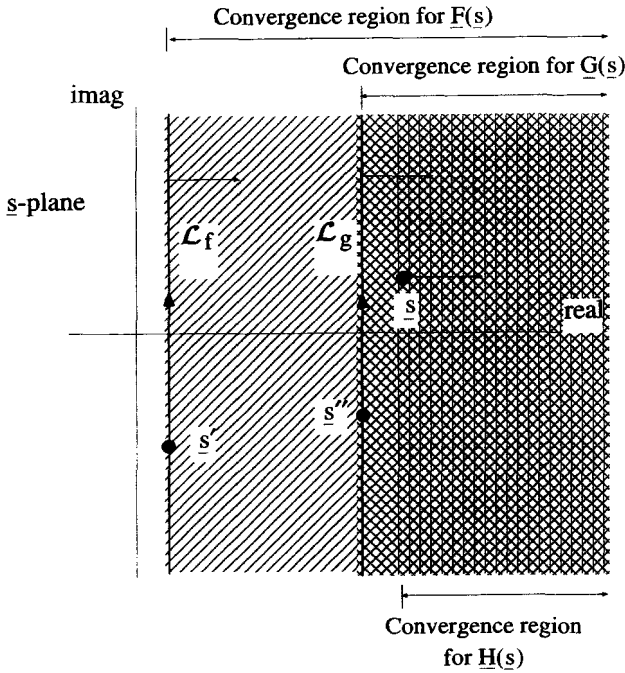
Figure 9.19   Convergence Regions for the Laplace Transform of a Product

this is a reasonable result, imagine $f(t)$ and $g(t)$ are both exponential functions. Their product is an exponential that is growing at a rate that is the sum of the growth rates of $f(t)$ and $g(t)$.

Along $\mathcal{L}_f$, $d\underline{s}' = ids_i'$, and on $\mathcal{L}_g$, $d\underline{s}'' = ids_i''$. Substituting these differentials into Equation 9.70 and using the result of Equation 9.72 gives:

$$\underline{H}(\underline{s}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} ds_i' \, \underline{F}(s_{ro}' + is_i') \int_{-\infty}^{\infty} ds_i'' \, \underline{G}(s_{ro}'' + is_i'') \, \delta(s_i' + s_i'' - s_i). \qquad (9.74)$$

These integrations can be performed in either order. Doing the $\int ds_i''$ first gives

$$\underline{H}(\underline{s}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} ds_i' \, \underline{F}(s_{ro}' + is_i') \, \underline{G}(s_{ro}'' + is_i - is_i'). \qquad (9.75)$$

Using the condition that $s_{ro}' + s_{ro}'' = s_r$, Equation 9.75 can be rewritten as

$$\underline{H}(\underline{s}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} ds_i' \, \underline{F}(s_{ro}' + is_i') \, \underline{G}(s_r + is_i - s_{ro}' - is_i'). \qquad (9.76)$$

This integration can now be converted back to one along the path $\mathcal{L}_f$:

$$\underline{H}(\underline{s}) = \frac{1}{2\pi i} \int_{\mathcal{L}_f} d\underline{s}' \, \underline{F}(\underline{s}') \, \underline{G}(\underline{s} - \underline{s}'). \qquad (9.77)$$

Notice the result of Equation 9.77 is similar to the convolution integral discussed earlier, except now the integration is done along a complex path. The result of Equation 9.77 can be written in a shorthand notation as

$$\mathcal{L}\{f(t)\,g(t)\} = \frac{1}{2\pi i}\left[\underline{F}(\underline{s}) \otimes_{\mathcal{L}_f} \underline{G}(\underline{s})\right]. \tag{9.78}$$

If the order of the integration of Equation 9.74 is reversed the result becomes

$$\mathcal{L}\left\{f(t)\,g(t)\right\} = \frac{1}{2\pi i}\left[\underline{G}(\underline{s}) \otimes_{\mathcal{L}_g} \underline{F}(\underline{s})\right]. \tag{9.79}$$

We have obtained a result that is similar to the one for Fourier transforms. The Laplace transform of the product of two functions is the convolution of the separate transforms. In this case, however, the convolution must be integrated on one of the Laplace contours.

In arriving at Equation 9.77 s_r was determined by the locations of the $\mathcal{L}_f$ and $\mathcal{L}_g$ contours according to the condition $s_r = s'_{ro} + s''_{ro}$. As the $\mathcal{L}_f$ and $\mathcal{L}_g$ contours are moved to the right, s_r increases and consequently $\underline{s}$ also moves to the right, as shown in Figure 9.19. When using Equations 9.78 or 9.79 to evaluate $\underline{H}(\underline{s})$, however, it is convenient to take another point of view. In particular, when using Equation 9.77, it makes more sense to select a fixed value for $\underline{s}$ and use this value to determine the location of the $\mathcal{L}_f$ contour. To do this, it must be realized that only values of $\underline{s}$ in the convergence region of $\underline{H}(\underline{s})$ can be used. In this convergence region, all the $\underline{s}'$-plane poles of $\underline{G}(\underline{s} - \underline{s}')$ must be to the right of all the $\underline{s}'$-plane poles of $\underline{F}(\underline{s}')$. This condition can be used to define the convergence region for $\underline{H}(\underline{s})$. The integration of Equation 9.77 is then evaluated by placing the $\mathcal{L}_f$ contour between the $\underline{s}'$-plane poles of $\underline{F}(\underline{s}')$ and $\underline{G}(\underline{s} - \underline{s}')$. These ideas are demonstrated in the following example.

Example 9.1 The process of convolution along a Laplace contour can be demonstrated by a simple example. Let

$$f(t) = \begin{cases} e^{2t} & t > 0 \\ 0 & t < 0 \end{cases}. \tag{9.80}$$

and

$$g(t) = \begin{cases} e^{3t} & t > 0 \\ 0 & t < 0 \end{cases}. \tag{9.81}$$

The Laplace transforms of these functions are easily taken:

$$\underline{F}(\underline{s}) = \frac{1}{\underline{s} - 2}, \tag{9.82}$$

with a convergence region given by $s_r > 2$, and

$$\underline{G}(\underline{s}) = \frac{1}{\underline{s} - 3}, \tag{9.83}$$

with a convergence region $s_r > 3$. If $h(t) = f(t)g(t)$ then

$$h(t) = \begin{cases} e^{5t} & t > 0 \\ 0 & t < 0 \end{cases}, \tag{9.84}$$

and its Laplace transform can be taken directly

$$\underline{H}(\underline{s}) = \frac{1}{\underline{s} - 5}, \tag{9.85}$$

with a convergence region $s_r > 5$.

This Laplace transform of $h(t)$ and its convergence region can also be obtained using Equation 9.77 and the convolution of $\underline{F}(\underline{s})$ with $\underline{G}(\underline{s})$:

$$\underline{H}(\underline{s}) = \frac{1}{2\pi i} \int_{\mathcal{L}_f} d\underline{s}' \left(\frac{1}{\underline{s}' - 2}\right) \left(\frac{1}{\underline{s} - \underline{s}' - 3}\right). \tag{9.86}$$

The integrand has two poles, one at $\underline{s}' = 2$ and another at $\underline{s}' = \underline{s} - 3$, as shown in Figure 9.20. The $\mathcal{L}_f$ contour must lie in between these poles as indicated in the figure. This contour can be closed to the left or to the right to produce the same result we obtained in Equation 9.85:

$$\underline{H}(\underline{s}) = \frac{1}{\underline{s} - 5}. \tag{9.87}$$

The convolution arguments also give the same region of convergence for this transform. The pole at $\underline{s} - 3$ must be to the right of the pole at 2, to make room for the $\mathcal{L}_f$ contour. Consequently, $s_r - 3 > 2$ and the convergence region for $\underline{H}(\underline{s})$ is given by $s_r > 5$.
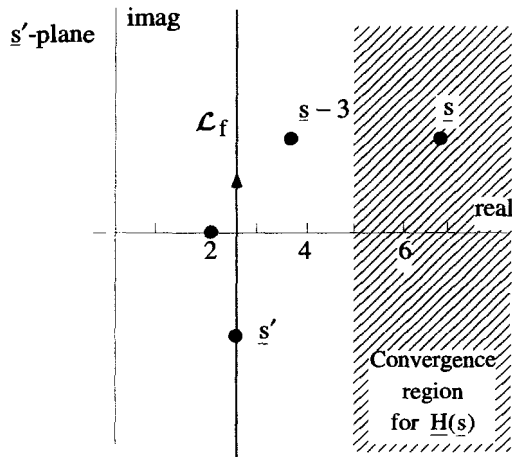

Figure 9.20   Convolution Integration in the Complex Plane

9.5.5 Summary of Transform Properties

Some properties of the Laplace transform operation are summarized in the table below. In this table the double arrow $\leftrightarrow$ is used to indicate the reversible transform process.

$$
\begin{array}{rcl}
f(t) & \leftrightarrow & \underline{F}(s) \\
g(t) & \leftrightarrow & \underline{G}(s) \\
f(t - t_o) & \leftrightarrow & e^{-st_o}\underline{F}(s) \\
df(t)/dt & \leftrightarrow & -f(0) + s\,\underline{F}(s) \\
f(t) \otimes g(t) & \leftrightarrow & \underline{F}(s)\,\underline{G}(s) \\
2\pi i\, f(t)\, g(t) & \leftrightarrow & \underline{F}(s)\,\otimes_{\mathcal{L}_f}\,\underline{G}(s)
\end{array}
$$

9.6 THE LAPLACE TRANSFORM CIRCUIT

As was the case with Fourier series and Fourier transforms, the inverse Laplace transform can be conceptually modeled with an additive electronic circuit. Consider the inverse Laplace transform integral

$$
f(t) = \frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s}\; e^{st}\underline{F}(\underline{s}), \tag{9.88}
$$

which is an integration along the contour $\mathcal{L}$, as shown in Figure 9.21. The Laplace contour must lie within the convergence region of the transform function which is to the right of all the poles of $\underline{F}(s)$. Along this contour, $\underline{s} = s_{ro} + is_i$, $d\underline{s} = ids_i$, and s_i ranges from $-\infty$ to ∞. With these facts in mind, Equation 9.88 can be rewritten as

$$
f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} ds_i\; e^{s_{ro}t}e^{is_it}\underline{F}(s_{ro} + is_i). \tag{9.89}
$$

If we think of the integral of Equation 9.89 in terms of its Riemann definition, it is the limit of a discrete sum of terms:

$$
f(t) = \lim_{\Delta s_i \to 0} \sum_{n=-\infty}^{\infty} \underline{F}(s_{ro} + in\Delta s_i)\frac{\Delta s_i}{2\pi}e^{s_{ro}t}e^{in\Delta s_i t}. \tag{9.90}
$$
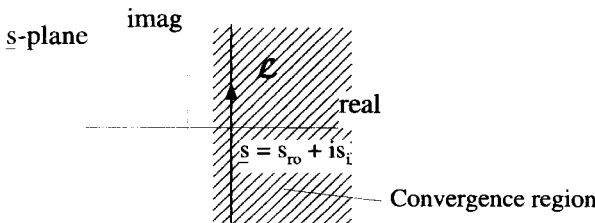
Figure 9.21 The Standard Laplace Contour

Equation 9.90 describes $f(t)$ as a sum of functions, each with an exponentially growing envelope of $(1/2\pi)\underline{F}(s_{ro} + in\Delta s_i)\Delta s_i e^{s_{ro}t}$, and oscillating as $e^{in\Delta s_i t}$.

9.6.1 A Growing Exponential

Consider the growing exponential

$$f(t) = \begin{cases} 0 & t < 0 \\ e^{\alpha t} & t > 0 \end{cases}, \tag{9.91}$$

with $\alpha > 0$. Its Laplace transform is given by

$$\underline{F}(\underline{s}) = \frac{1}{\underline{s} - \alpha}. \tag{9.92}$$

The summation on the RHS of Equation 9.90 can be modeled with the circuit shown in Figure 9.22. There is a fair amount of information in this figure. The insert in the upper right-hand corner shows the $\underline{s}$-plane with its pole, the convergence region
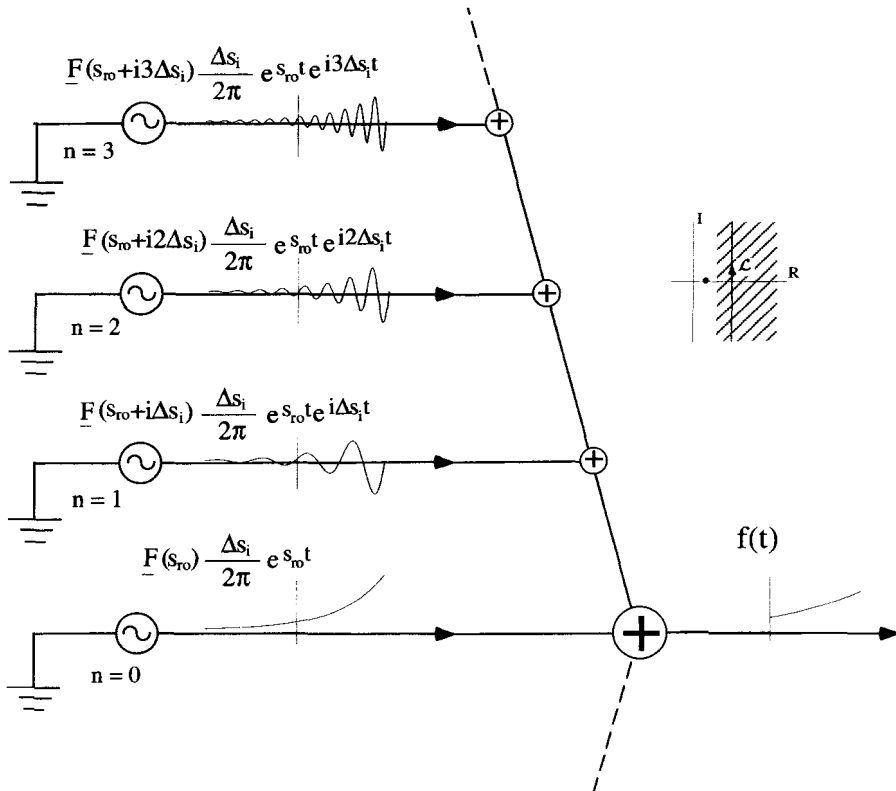
Figure 9.22　The Basic Laplace Circuit
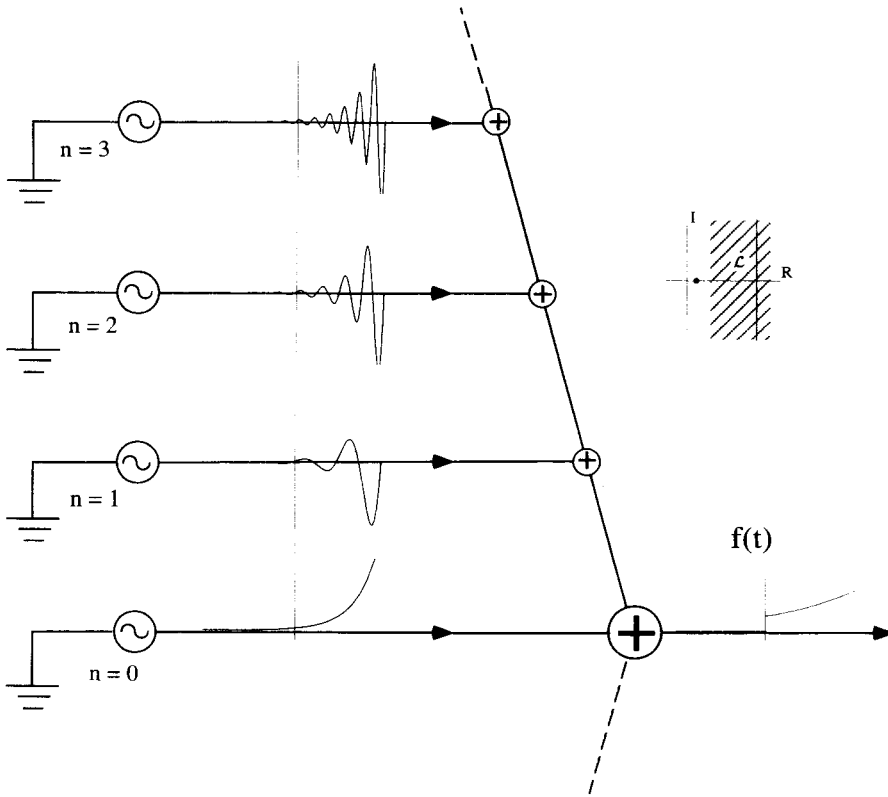
Figure 9.23   The Laplace Circuit with the Laplace Contour Shifted to the Right

of the transform, and the inversion contour. Each term in the series on the RHS of Equation 9.90 is represented by an oscillator. In this figure only the first four oscillators for $n \geq 0$ have been shown. All the oscillators have been going on for all time and continue to go on for all time. When added together they generate $f(t)$, a function that is zero for all $t < 0$, and grows exponentially for $t > 0$. The exponential growth rate of $f(t)$ must be less than $e^{s_{ro}t}$, the growth rate of the individual oscillators.

The $\mathcal{L}$ contour can be moved to the right without affecting the result of the integration in Equation 9.89. This means we can increase s_{ro}, making each oscillator have an arbitrarily large growth rate. Nevertheless, when they are all added up, they generate the same $f(t)$. This somewhat surprising result is depicted in Figure 9.23.

9.6.2   A Decaying Exponential

Things get even more interesting when we consider the case where the single pole of $\underline{F}(\underline{s})$ lies on the negative real axis. This happens in Equation 9.91 when $\alpha < 0$. The region of convergence for the Laplace transform integration still lies to the right of the pole at $\underline{s} = \alpha$, and the inversion contour $\mathcal{L}$ can be placed anywhere to the right
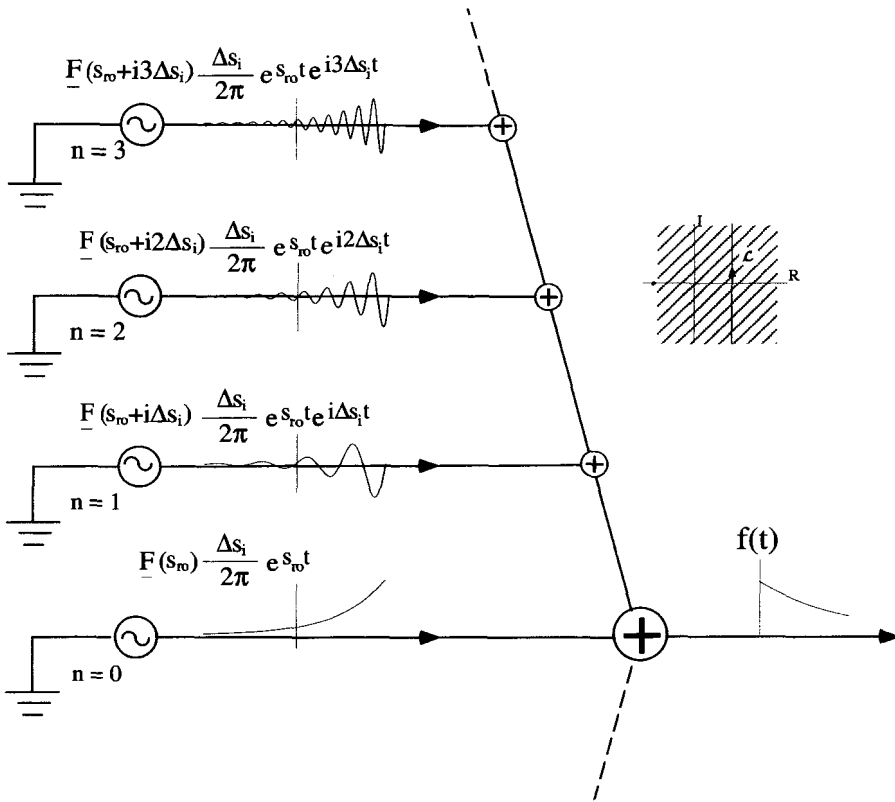
Figure 9.24   The Laplace Circuit for an Exponentially Damped Function

of this pole. Let's start with it in the right half plane so that $s_{ro} > 0$. The circuit for this situation is similar to the previous ones and is shown in Figure 9.24. The inserted figure shows the Laplace contour and region of convergence. The sum of function generators produces an $f(t)$ that is zero for $t < 0$. For $t > 0$, even though each function generator produces a signal that diverges as t increases, the sum is a damped exponential. Think about that last sentence for a second. We are adding up an infinite number of exponentially growing, oscillating functions which miraculously add in just the right way to generate a smooth, decaying exponential. Very strange, indeed.

Because the convergence region of this transform extends into the left half plane, we can move $\mathcal{L}$ to the left, until it sits on top of the imaginary axis with $s_{ro} = 0$. In this case, the Laplace inversion has effectively gone to a Fourier inversion and $\underline{F}(is_i)$ is really just the Fourier transform of $f(t)$. The Laplace circuit is now composed of signal generators oscillating at a constant amplitude for all time, as shown in Figure 8.6.

Let's keep shifting $\mathcal{L}$ to the left, so that it lies in the left half-plane with $s_{ro} < 0$. The Laplace circuit for this case is shown in Figure 9.25. Now the circuit is composed of signal generators whose amplitudes are decaying, because s_{ro} is less than zero. For
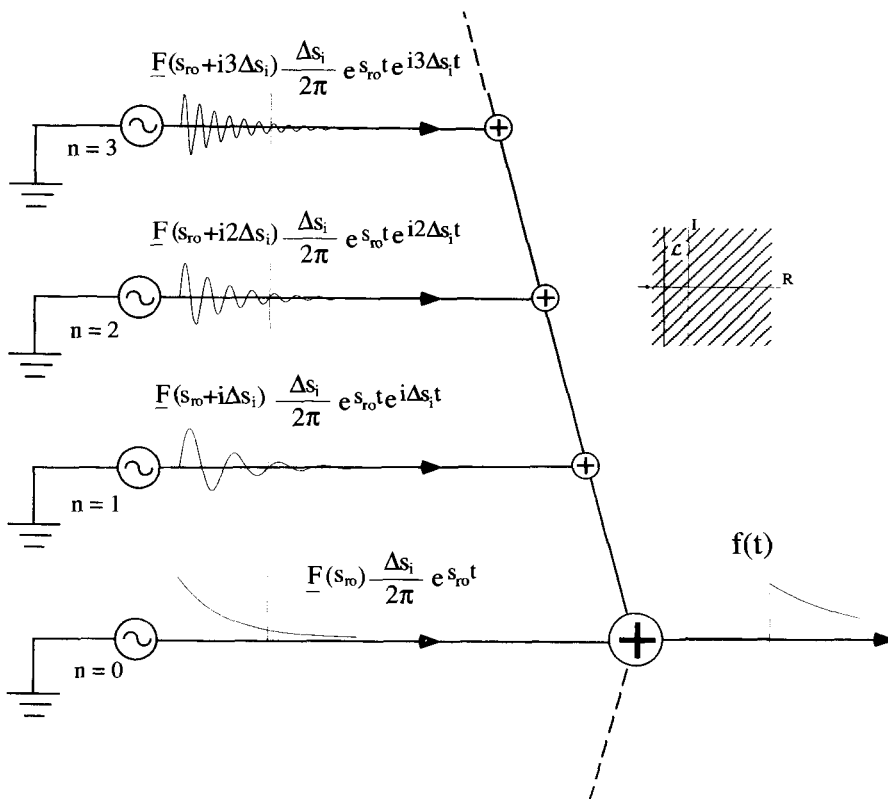
Figure 9.25 The Laplace Circuit with the Laplace Contour in the Left Half $\underline{s}$-Plane

$t > 0$ the signals from the generators sum to give the damped exponential of $f(t)$. Its damping rate α is greater than the damping rate of the individual signal generators because $\alpha < s_{ro}$. For $t < 0$, even though the amplitudes of the individual oscillators are diverging as t goes to $-\infty$, their sum produces a signal that is zero for all $t < 0$.

9.7 DOUBLE-SIDED OR BILATERAL LAPLACE TRANSFORMS

Early on, we made the statement that the Laplace transform is more general than the Fourier transform and can handle any function that the Fourier transform operation can. However, for the regular Laplace transform, $f(t)$ must be zero for $t < 0$. The function

$$f(t) = \begin{cases} e^{-\alpha_1 t} & t > 0 \\ e^{+\alpha_2 t} & t < 0 \end{cases}, \qquad (9.93)$$

where α_1 and α_2 are both positive, real numbers, is absolutely integrable and clearly possesses a legitimate Fourier transform. But it is not zero for $t < 0$, and therefore

does not have a regular Laplace transform. So how can we make the claim that any function which can be Fourier transformed can also be Laplace transformed?

It is true that the regular Laplace transform, with the restriction that $f(t)$ must be zero for $t < 0$, cannot handle a function like the one in Equation 9.93. However, a small modification to the regular Laplace transform clears up this problem.

Consider the Laplace transform *type* operation

$$\underline{F}(\underline{s}) = \int_{-\infty}^{\infty} dt \, e^{-\underline{s}t} f(t), \tag{9.94}$$

where $t < 0$ is now included in the integration, and $f(t)$ is given by Equation 9.93. The obvious question immediately arises. Is there is a region of convergence in the $\underline{s}$-plane where the integral of Equation 9.94 exists? Inserting Equation 9.93 into 9.94 gives

$$\underline{F}(\underline{s}) = \int_{-\infty}^{0} dt \, e^{-\underline{s}t} e^{\alpha_2 t} + \int_{0}^{\infty} dt \, e^{-\underline{s}t} e^{-\alpha_1 t}$$

$$= \frac{e^{(\alpha_2 - \underline{s})t}}{\alpha_2 - \underline{s}} \Big|_{-\infty}^{0} + \frac{e^{-(\alpha_1 + \underline{s})t}}{-\alpha_1 - \underline{s}} \Big|_{0}^{\infty}. \tag{9.95}$$

The first term on the RHS of Equation 9.95 converges for $s_r < \alpha_2$, and the second term converges for $s_r > -\alpha_1$. Therefore

$$\underline{F}(\underline{s}) = -\frac{\alpha_1 + \alpha_2}{(\underline{s} + \alpha_1)(\underline{s} - \alpha_2)}, \tag{9.96}$$

with a convergence region given by $-\alpha_1 < s_r < \alpha_2$. The function $\underline{F}(\underline{s})$ has first-order poles at $\underline{s} = \alpha_2$ and $\underline{s} = -\alpha_1$ (remember that both α_1 and α_2 are positive, real numbers). This convergence region is therefore a strip between the poles of $\underline{F}(\underline{s})$, as shown in Figure 9.26.

It seems reasonable to try to set up the inversion integral along a Laplace contour in the convergence region as shown in Figure 9.26. This inversion integral becomes

$$\frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s} \, \frac{-e^{\underline{s}t}(\alpha_1 + \alpha_2)}{(\underline{s} + \alpha_1)(\underline{s} - \alpha_2)}. \tag{9.97}$$
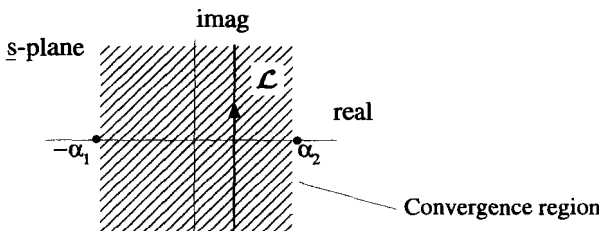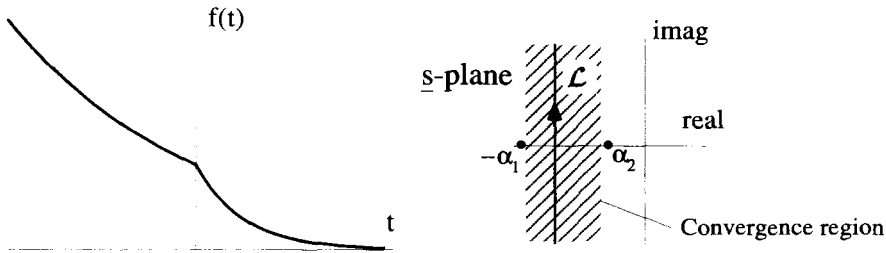

Figure 9.26   Convergence Region for a Double-Sided Laplace Transform

Figure 9.27   The Convergence Region for a Function that Grows Exponentially for $t < 0$

Notice, there are poles of $\underline{F}(\underline{s})$ on either side of $\mathcal{L}$. In order to close this contour, we have the same requirements as we did for the regular Laplace transform. Because of the $e^{\underline{s}t}$ term, we must close to the left for $t > 0$ and to the right for $t < 0$. After evaluating the residues of the poles, it is quite easy to see the contour integral of Equation 9.97 generates the original $f(t)$ in Equation 9.93. The Laplace contour can be placed anywhere in the convergence region with the same result. If it is placed on the imaginary $\underline{s}$-axis, this double-sided Laplace transform operation is equivalent, except for a $\sqrt{2\pi}$ factor, to a Fourier transform.

If we let α_1 or α_2 become negative, the standard Fourier transform will fail. But, as long as $\alpha_2 > -\alpha_1$, there will be a finite convergence region and this special Laplace transform will exist. In cases like this, the convergence region does not include the imaginary $\underline{s}$-axis, consistent with the fact that the Fourier transform operation fails. An example of this, with $\alpha_2 < 0$, is shown in Figure 9.27.

Therefore, this more general approach to the Laplace transform, which is called a double-sided or bilateral Laplace transform, includes all the properties of the Fourier transform. To use this approach we must take more care specifying the Laplace contour and the region of convergence. This is important, as you will see in the example below, because a given $\underline{F}(\underline{s})$ can generate different $f(t)$'s depending on how the convergence region is defined.

Example 9.2   Consider the transform given by Equation 9.96 with the specific values $\alpha_1 = \alpha_2 = 1$:

$$\underline{F}(\underline{s}) = -\frac{2}{(\underline{s} + 1)(\underline{s} - 1)}. \tag{9.98}$$

This function has simple poles at $\underline{s} = +1$ and $\underline{s} = -1$. There are three possible regions of convergence: $s_r > 1$; $-1 < s_r < 1$; and $s_r < -1$. Laplace inversions using these three different convergence regions result in three different functions of t.

For the first case, with the region of convergence given by $s_r > 1$, the Laplace contour is to the right of all the poles, as shown in Figure 9.28. This is just a regular Laplace transform. Because there are no poles to the right of the Laplace contour $f(t)$
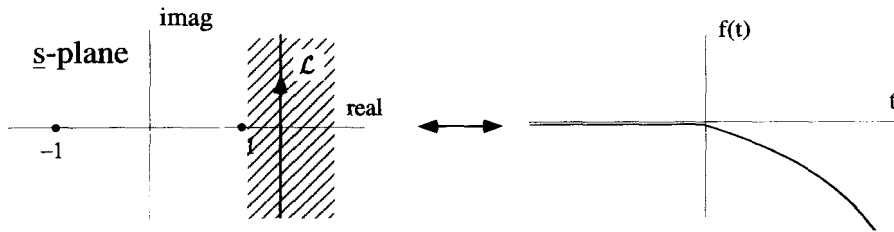
Figure 9.28 Inversion for the Region of Convergence Given by Real $\underline{s} > 1$

is zero for $t < 0$. For $t > 0$, $f(t)$ is given by the sum of the residues of the two poles:

$$f(t) = \begin{cases} 0 & t < 0 \\ e^{-t} - e^t & t > 0 \end{cases}, \tag{9.99}$$

as shown in Figure 9.28. This $f(t)$ has no Fourier transform analogue, because the region of convergence does not include the imaginary $\underline{s}$-axis.

The second region of convergence, given by $-1 < s_r < 1$, is the region between the two poles, as shown in Figure 9.29. This is a specific case of the situation we discussed previously. The Laplace contour lies between the two poles, and now $f(t)$ is no longer zero for $t < 0$, but is given by minus the residue of the pole at $\underline{s} = 1$. Similarly, $f(t)$ for $t > 0$ is nonzero and equals the residue of the pole at $\underline{s} = -1$,

$$f(t) = \begin{cases} e^{+t} & t < 0 \\ e^{-t} & t > 0 \end{cases}, \tag{9.100}$$

as shown in Figure 9.29. In this case the imaginary $\underline{s}$-axis is in the region of convergence, so there is a corresponding Fourier transform.

The last region of convergence has $s_r < -1$, and is shown in Figure 9.30. In this case, the contour lies to the left of all the poles, and consequently $f(t)$ is zero for $t > 0$. For $t < 0$, we sum the negative of the residues of the poles to get,

$$f(t) = \begin{cases} e^t - e^{-t} & t < 0 \\ 0 & t > 0 \end{cases}, \tag{9.101}$$
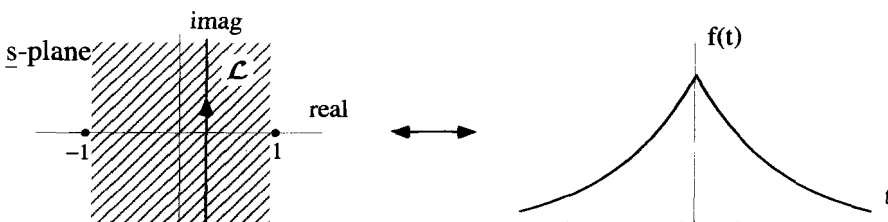
Figure 9.29 Inversion for the Convergence Region Given by $-1 < $ Real $\underline{s} < 1$
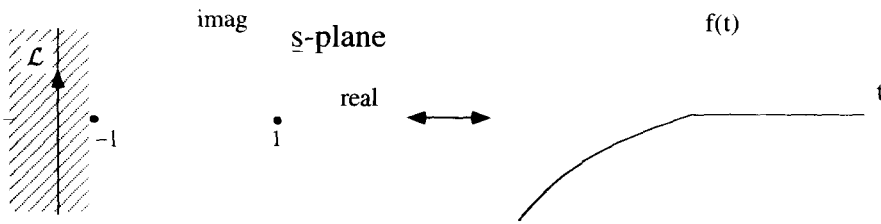
Figure 9.30   Inversion for the Convergence Region Given by $-1 <$ Real $\underline{s}$

as shown in Figure 9.30. This function does not possess a Fourier transform, because the imaginary $\underline{s}$-axis is not in the region of convergence.

EXERCISES FOR CHAPTER 9

1. Find the Laplace transform of the function

$$f(t) = \begin{cases} e^{\alpha_o t} \ \sin(\omega_o t) & t > 0 \\ 0 & t < 0 \end{cases},$$

where α_o and ω_o are positive real numbers. Demonstrate that you have obtained the correct transform by inverting it to obtain the original $f(t)$.

2. Consider the function

$$f(t) = \begin{cases} e^t & t > -1 \\ 0 & t < -1 \end{cases}.$$

Sketch $f(t)$ vs. t and determine its Laplace transform and region of convergence. Confirm that you have the right transform by inverting it to obtain the original $f(t)$.

3. Show that the Laplace transform of

$$f(t) = \begin{cases} t \ \sin(kt) & t > 0 \\ 0 & t < 0 \end{cases}$$

is

$$\underline{F}(\underline{s}) = \frac{2k\underline{s}}{(\underline{s}^2 + k^2)^2}.$$

Determine the region of convergence for $\underline{F}(\underline{s})$.

4. Find the Laplace transform of the function

$$f(t) = \begin{cases} t^3 e^t & t > 0 \\ 0 & t < 0 \end{cases}$$

(a) using the standard transform integral and integrating by parts the appropriate number of times.

(b) by treating the function as a product of two other functions.

5. Determine the Laplace transforms of the following functions. Assume the functions are zero for $t < 0$, and for $t > 0$ are given by:

(a) $\sinh(kt)$.

(b) $\cosh(kt)$.

(c) $\sin(kt)$.

(d) $\cos(kt)$.

(e) $e^t \cos t$

(f) t^n.

6. Find the inverse Laplace transform of

$$\underline{F}(s) = \frac{s}{(s^2 + 1)^2},$$

with the region of convergence $s_r > 0$. Make a sketch of $f(t)$.

7. Find the function that has the Laplace transform

$$\underline{F}(s) = \frac{1}{(s - 1)^2},$$

with the convergence region $s_r > 1$.

8. Find the inverse transform of

$$\underline{F}(s) = \frac{1}{s^2(s - k)},$$

where k is a real positive number

(a) by direct integration of $\underline{F}(s)$.

(b) by factoring $\underline{F}(s)$ and using the convolution theorem.

(c) by expanding $\underline{F}(s)$ as a sum of partial fractions.

9. Find the inverse Laplace transform of

$$\underline{F}(s) = \frac{1}{s(s - k)^2},$$

where k is a positive real number

(a) by direct integration of $\underline{F}(s)$.

(b) by factoring $\underline{F}(s)$ and using the convolution theorem.

(c) by expanding $\underline{F}(s)$ as a sum of partial fractions.

10. Find the inverse Laplace transform of

$$\underline{F}(\underline{s}) = \frac{k^2}{\underline{s}(\underline{s}^2 + k^2)}$$

   (a) by direct integration of $\underline{F}(\underline{s})$.
   (b) by factoring $\underline{F}(\underline{s})$ and using the convolution theorem.
   (c) by expanding $\underline{F}(\underline{s})$ as a sum of partial fractions.

11. Show that

$$\mathcal{L}\left[\frac{d^2 f(t)}{dt^2}\right] = \underline{s}^2 \mathcal{L}[f(t)] - \underline{s}\, f(0) - \left.\frac{df(t)}{dt}\right|_{t=0},$$

where $\mathcal{L}[f(t)] = \underline{F}(\underline{s})$. Develop a general expression for

$$\mathcal{L}\left[\frac{d^n f(t)}{dt^n}\right].$$

12. If

$$\underline{G}(\underline{s}) = \frac{d\underline{F}(\underline{s})}{d\underline{s}},$$

what is $g(t)$, the inverse transform of $\underline{G}(\underline{s})$, in terms of $f(t)$, the inverse transform of $\underline{F}(\underline{s})$?

13. If the function

$$\underline{F}(\underline{s}) = \frac{e^{-\underline{s}}}{\underline{s}^2 - 1}$$

is treated as a bilateral Laplace transform, it can have three different regions of convergence:
   (a) $s_r < -1$.
   (b) $-1 < s_r < 1$.
   (c) $1 < s_r$.
For each of these convergence regions, the inversion of $\underline{F}(\underline{s})$ gives a different function of time. Determine these three functions and make a plot of them.

14. Find the Laplace transform of

$$f(t) = \begin{cases} 0 & t > -1 \\ e^{-t} & t < -1 \end{cases}.$$

15. Find the inverse Laplace transform of

$$\underline{F}(\underline{s}) = \frac{\underline{s}}{(\underline{s}^2 - 1)^2},$$

assuming that the region of convergence is $-1 < s_r < +1$.

16. Find the Laplace transform of $f(t)$ where

$$f(t) = \begin{cases} 0 & t > 0 \\ \sin(\omega_o t) & t < 0 \end{cases}$$

and identify its region of convergence in the $\underline{s}$-plane. Invert this transform using closure to obtain the original $f(t)$.

17. Consider the function

$$f(t) = \begin{cases} e^{-\alpha t} & t > 0 \\ e^{-\beta t} & t < 0 \end{cases},$$

where α and β are both positive, real numbers.

(a) Make a sketch of this function.

(b) Why does this function not have a standard Laplace transform?

(c) Define a transform pair that will work for this function. Discuss the conditions on α and β that allow a finite region of convergence for the transform function.

10

DIFFERENTIAL EQUATIONS

A differential equation contains terms involving either total (d/dx) or partial $(\partial/\partial x)$ differential operations. In this chapter, several methods for obtaining solutions to differential equations will be presented. The emphasis is on solutions to linear problems, although some techniques for solving nonlinear equations will be discussed. Particular attention will be paid to Fourier and Laplace transform approaches and Green's function solutions. A basic knowledge of differential equations is assumed.

10.1 TERMINOLOGY

The study of differential equations involves a large number of definitions. This section presents a quick review of some of these definitions and other important terminology.

10.1.1 Dependent versus Independent Variables

Two types of variables appear in differential equations. A dependent variable represents the unknown solution to the equation and typically appears in the numerator of the differential operation(s). An independent variable is a parameter of the solution. The dependent variables are functions of the independent variables. An independent variable typically appears in the denominator of the differential operation(s). For example, in the differential equation

$$\frac{dy(x)}{dx} = y(x) + 5, \tag{10.1}$$

x is an independent variable and y is a dependent variable. The dependence of y on x is indicated by writing $y(x)$.

10.1.2   Ordinary versus Partial Differential Equations

Ordinary differential equations involve only one independent variable and contain only total differential operations. For example,

$$\frac{dy(x)}{dx} + y(x) = 0 \tag{10.2}$$

is a typical, ordinary differential equation. There is a single dependent variable y, which depends on the single independent variable x.

On the other hand, in a partial differential equation the dependent variable is a function of more than one independent variable. Consequently, these equations contain partial differential operations. For example,

$$\frac{\partial y(x,t)}{\partial x} - 5\frac{\partial y(x,t)}{\partial t} + y(x,t) = 0 \tag{10.3}$$

is a partial differential equation. It involves one dependent variable y, which depends on the two independent variables x and t. This dependence is emphasized by writing $y(x,t)$.

10.1.3   Linear versus Nonlinear Equations

A linear differential equation contains only terms that are linear in the dependent variables or their derivatives. In a linear equation, if y is a dependent variable, there can be no y^2, $y(dy/dt)$, $(dy/dt)^3$, etc. terms, nor terms where the dependent variable appears in the argument of a nonlinear function. There are no restrictions on the independent variables. Equations 10.2 and 10.3 are examples of linear differential equations. The equation

$$\frac{dy(x)}{dx} + x^2 y(x) = \sin(x) \tag{10.4}$$

is also a linear, even though it contains an x^2 factor and a $\sin(x)$ term.

A nonlinear differential equation contains dependent variable products or nonlinear functions of the dependent variable. For example,

$$\frac{d\left[y^2(x)\right]}{dx} + \left[\frac{dy(x)}{dx}\right]^2 + x^2 y(x) = 0 \tag{10.5}$$

and

$$\frac{dy(x)}{dx} + x^2 \sin(y) = 0 \tag{10.6}$$

are nonlinear differential equations. The first is nonlinear because $y(x)$ is squared in the numerator of the differential operation and because of the $(dy/dx)^2$ term. The second is nonlinear because the dependent variable appears inside the nonlinear sine operation.

Except in rare cases, nonlinear differential equations are much more difficult to solve than linear ones. Many times nonlinear equations require numerical or approximate solutions.

10.1.4 Order

The order of a differential equation, regardless of whether it involves partial or total differential operations, is determined by the highest derivative present in the equation. Equations 10.2 and 10.3 are examples of first-order differential equations, and

$$\frac{d^3 y(x)}{dx^3} + \frac{dy(x)}{dx} - y(x) = 0 \tag{10.7}$$

is an example of a third-order differential equation.

10.1.5 Coupled Differential Equations

Coupled differential equations involve more than one dependent variable. For example,

$$\frac{d^2 x_1(t)}{dt^2} = x_1(t) - x_2(t) \tag{10.8}$$

$$\frac{d^2 x_2(t)}{dt^2} = x_2(t) - x_1(t) \tag{10.9}$$

form a set of coupled, second-order, linear differential equations. Both x_1 and x_2 are dependent variables and are functions of the single independent variable t.

10.1.6 Homogeneous versus Nonhomogeneous Equations

We will restrict our comments on homogeneous and nonhomogeneous properties to linear equations. A linear differential equation is homogeneous if every term contains the dependent variable. Homogeneous equations are usually arranged so that all the terms are on the LHS. Equation 10.7 is one example of a homogeneous equation. In general, an ordinary homogeneous differential equation can always be placed in the form

$$\mathcal{L}_{op} y(x) = 0, \tag{10.10}$$

where $\mathcal{L}_{op}$ is a linear differential operator, which acts on the dependent variable y. The operator for Equation 10.7 would be

$$\mathcal{L}_{op} = \frac{d^3}{dx^3} + \frac{d}{dx} - 1. \tag{10.11}$$

Solutions to homogeneous equations are sometimes called "undriven" solutions, because they can describe the response of a system which is not under the influence of an external force. The solutions of homogeneous equations are not unique, but can always be scaled by an arbitrary constant. A homogeneous equation will always have a trivial solution of $y(x) = 0$.

The superposition principle applies to solutions of linear, homogeneous equations. It states that if $y_1(x)$ and $y_2(x)$ are solutions to a linear, homogeneous differential

equation, then a linear sum of these solutions, $y(x) = A_o\,y_1(x) + B_o\,y_2(x)$ where A_o and B_o are constants, is also a solution to the same differential equation.

The order of a linear, homogeneous differential equation is especially important. It is equal to the number of nontrivial (nonzero) solutions to the equation, and consequently to the number of boundary conditions necessary to make the solution unique.

In contrast, nonhomogeneous differential equations contain at least one term without a dependent variable. They are sometimes referred to as "driven" equations, because they describe the response of a system to an external signal. Nonhomogeneous equations are usually arranged so that the terms that contain the dependent variable(s) are all on the left, and all other terms are on the right. Using the operator notation, an ordinary, linear, nonhomogeneous equation can be written

$$\mathcal{L}_{op}y(x) = f(x), \tag{10.12}$$

where $f(x)$ is often called the driving function. The equation

$$\frac{d^2\phi(x)}{dx^2} + \phi(x) = \sin x \tag{10.13}$$

is an example of a linear, nonhomogeneous equation.

The general solution to a linear nonhomogeneous differential equation $\mathcal{L}_{op}y(x) = f(x)$ can always be written as the sum of two parts

$$y(x) = y_h(x) + y_p(x). \tag{10.14}$$

The first term, called the homogeneous part, is the general solution of the corresponding homogeneous equation $\mathcal{L}_{op}y(x) = 0$. The second part, called the particular solution, is any solution of the nonhomogeneous equation.

10.2   SOLUTIONS FOR FIRST-ORDER EQUATIONS

Many first-order equations are relatively easy to solve because the solutions can be obtained simply by integrating the differential equations once. However, there are some first-order equations that are not immediately in integrable form. In this section, three standard approaches for these types of problems are presented: the separation of variables, the formation of an exact differential, and the use of an integrating factor.

10.2.1   Separation of Variables

Any first-order differential equation that can be placed in the form

$$\frac{dy(x)}{dx} = -\frac{P(x)}{Q(y)} \tag{10.15}$$

may be able to be solved using the method of separation of variables. The minus sign in this equation is an arbitrary convention. The differential equation may be nonlinear through the $Q(y)$ term.

Rewrite Equation 10.15 as

$$Q(y)\, dy = -P(x)\, dx. \tag{10.16}$$

Now if both sides of Equation 10.16 can be integrated, we can generate a functional relationship between x and y, with one arbitrary constant of integration. In physical problems, that constant is determined by one boundary condition that must be specified to make the solution unique.

Example 10.1   The nonlinear differential equation

$$\frac{dy}{dx} = \frac{\cos(x)}{y}, \tag{10.17}$$

with the boundary condition of $y(0) = 0$, can be solved by separation of variables. Equation 10.17 leads to the integral equation

$$\int dy\, y = \int dx\, \cos(x). \tag{10.18}$$

Both sides of Equation 10.18 easily integrate to give

$$\frac{y^2}{2} + C_o = \sin(x), \tag{10.19}$$

where the two arbitrary constants of integration have been combined into the single constant C_o. Imposing the boundary condition $y(0) = 0$ forces $C_o = 0$. Thus we have the final solution

$$y = \sqrt{2\sin(x)}. \tag{10.20}$$

10.2.2   Forming an Exact Differential

If it is not possible to separate variables, an approach that sometimes works is to form an exact differential. Try to set up the equation in the form

$$\frac{dy(x)}{dx} = -\frac{R(x, y)}{S(x, y)}. \tag{10.21}$$

Equation 10.21 can immediately be manipulated to read

$$R(x, y)\, dx + S(x, y)\, dy = 0. \tag{10.22}$$

This equation is in the form of an exact differential if there is a function $\phi(x, y)$ such that

$$\frac{\partial \phi(x, y)}{\partial x} = R(x, y) \tag{10.23}$$

and

$$\frac{\partial \phi(x, y)}{\partial y} = S(x, y). \tag{10.24}$$

If such a $\phi(x, y)$ exists, the LHS of Equation 10.22 can be written as

$$R(x, y) \, dx + S(x, y) \, dy = \frac{\partial \phi(x, y)}{\partial x} \, dx + \frac{\partial \phi(x, y)}{\partial y} \, dy$$

$$= d\phi, \tag{10.25}$$

and Equation 10.22 becomes

$$d\phi = 0. \tag{10.26}$$

The solution to Equation 10.21, the original differential equation, is therefore $\phi(x, y) = C_o$, where C_o is some constant. As before, this constant is determined by the boundary condition for the problem.

There is a simple test to see if a pair of functions forms an exact differential. If $R(x, y)$ and $S(x, y)$ form an exact differential, Equations 10.23 and 10.24 must be satisfied. If x and y are independent variables and the partial derivatives of $\phi(x, y)$ exist, then

$$\frac{\partial^2 \phi(x, y)}{\partial x \partial y} = \frac{\partial^2 \phi(x, y)}{\partial y \partial x}. \tag{10.27}$$

This means that for $R(x, y)$ and $S(x, y)$ to form an exact differential, they must satisfy the condition

$$\frac{\partial R(x, y)}{\partial y} = \frac{\partial S(x, y)}{\partial x}. \tag{10.28}$$

Example 10.2   Consider the differential equation

$$\frac{dy}{dx} = -\frac{3x^2 + 2y^2}{4xy}, \tag{10.29}$$

with the boundary condition $y(1) = 1$. Applying the test given by Equation 10.28 to the functions on the RHS of Equation 10.29 shows that an exact differential may be found for this problem. Indeed, the function

$$\phi(x, y) = x^3 + 2xy^2 \tag{10.30}$$

generates the two functions

$$R(x, y) = \frac{\partial \phi(x, y)}{\partial x} = 3x^2 + 2y^2$$

$$S(x, y) = \frac{\partial \phi(x, y)}{\partial y} = 4xy. \tag{10.31}$$

The solution to Equation 10.29 is therefore given by

$$x^3 + 2xy^2 = C_o. \tag{10.32}$$

The boundary condition $y(1) = 1$ determines $C_o = 3$, and our final result is

$$y = \sqrt{\frac{3 - x^3}{2x}}. \tag{10.33}$$

10.2.3 Integrating Factors

Integrating factors can sometimes be used to solve linear, nonhomogeneous, first-order differential equations of the form

$$\frac{dy(x)}{dx} + P(x)\,y(x) = G(x). \tag{10.34}$$

In this equation, $P(x)$ and $G(x)$ are known functions of the independent variable x. An integrating factor is a function $\alpha(x)$, by which we multiply both sides of Equation 10.34, so the LHS is placed in the form

$$\frac{d}{dx}[\alpha(x)\,y(x)].$$

In other words, we seek an $\alpha(x)$ such that

$$\alpha(x)\left[\frac{dy(x)}{dx} + P(x)\,y(x)\right] = \frac{d}{dx}[\alpha(x)\,y(x)]. \tag{10.35}$$

Using the product rule of differentiation on the RHS of Equation 10.35 gives a simple first-order differential equation for $\alpha(x)$:

$$\alpha(x)\,P(x) = \frac{d\alpha(x)}{dx}. \tag{10.36}$$

This equation can be rearranged as

$$\frac{d\alpha}{\alpha} = P(x)\,dx \tag{10.37}$$

and solved for $\alpha(x)$ as follows:

$$\int \frac{d\alpha}{\alpha} = \int dx\,P(x)$$

$$\ln \alpha = \int dx\,P(x) \tag{10.38}$$

$$\alpha(x) = e^{\int dx\,P(x)}.$$

Notice, it must be possible to calculate the integral of $P(x)$ in order to use this method. It is not necessary to worry about the constants of integration associated with the two integrations performed in the steps above. The original differential equation we are trying to solve is first order, and therefore needs only one arbitrary constant. This constant will be introduced in the subsequent steps.

With $\alpha(x)$ determined, the solution to the original differential equation can now be obtained. Multiplying both sides of Equation 10.34 by $\alpha(x)$ and using Equation 10.35 gives

$$\frac{d}{dx}[\alpha(x) y(x)] = \alpha(x) G(x).$$ (10.39)

Integrating both sides results in the expression

$$\alpha(x)y(x) = \int dx\, \alpha(x)G(x) + C_o.$$ (10.40)

The integration constant C_o is determined by the single boundary condition that is necessary to make the solution unique. The final result is

$$y(x) = \frac{\int dx\, \alpha(x)G(x) + C_o}{\alpha(x)}.$$ (10.41)

Notice that to use this process, not only must we be able to integrate the function $P(x)$ to get $\alpha(x)$, but we must also be able to integrate the function $\alpha(x)G(x)$.

Example 10.3   As a simple example of using an integrating factor, consider the differential equation

$$\frac{dy}{dx} + y = e^x,$$ (10.42)

with the boundary condition $y(0) = 1$. The integration factor is simply

$$a(x) = e^{\int dx} = e^x.$$ (10.43)

Substitution into Equation 10.41 gives

$$y(x) = \frac{\int dx\, e^{2x} + C_o}{e^x}$$

$$= \frac{1}{2}e^x + C_o e^{-x}.$$ (10.44)

The boundary condition requires $C_o = 1/2$, so

$$y(x) = \cosh(x).$$ (10.45)

10.3 TECHNIQUES FOR SECOND-ORDER EQUATIONS

Second-order differential equations pervade physics and engineering problems. One of the most common equation in physics is the second-order equation

$$\frac{d^2 y}{dx} = -\omega_o^2 y, \tag{10.46}$$

which describes the motion of a simple harmonic oscillator with a period of $2\pi/\omega_o$. As mentioned earlier, the number of nontrivial, independent solutions of a linear differential equation is equal to the order of the equation. Thus, a linear second-order equation should have two different solutions. We can see that this is the case for Equation 10.46, since both

$$y(x) = \sin(\omega_o x) \tag{10.47}$$

$$y(x) = \cos(\omega_o x) \tag{10.48}$$

are valid solutions.

There is no general analytic method for finding solutions to second-order and higher equations. Solutions are often obtained by educated guessing, or numerical methods. There is a general technique, however, for a specialized class of equations which have constant coefficients. In this section, we describe a simple technique for solving ordinary, second-order, linear differential equations with constant co-efficients. Then we introduce an important quantity called the Wronskian, which provides a powerful method for generating a second solution to a homogeneous differential equation when one solution is known. The Wronskian can also help generate the particular solution to a nonhomogeneous equation, if the general homogeneous solution is already known.

10.3.1 Ordinary, Homogeneous Equations with Constant Coefficients

The most general form of a second-order, linear, ordinary, homogeneous differential equation can be written as

$$\frac{d^2 y(x)}{dx^2} + P(x)\frac{dy(x)}{dx} + Q(x)\,y(x) = 0. \tag{10.49}$$

Equation 10.49 is said to have constant coefficients, if the functions $P(x)$ and $Q(x)$ are constants,

$$\frac{d^2 y(x)}{dx^2} - 2p\frac{dy(x)}{dx} + q\,y(x) = 0, \tag{10.50}$$

where the factor of -2 was used in the second term to simplify the algebra to come.

The manipulation of the linear equations is assisted by defining the differential operator of order n as $\mathcal{D}_{op}^n \equiv d^n/dx^n$. In this notation, Equation 10.50 becomes

$$(\mathcal{D}_{op}^2 - 2p\mathcal{D}_{op} + q)y(x) = 0. \tag{10.51}$$

We can now perform a fancy trick, and factor this as if it were a quadratic algebraic expression,

$$(\mathcal{D}_{op} - r_+)\underbrace{(\mathcal{D}_{op} - r_-)y(x)}_{s(x)} = 0, \qquad (10.52)$$

where $r_\pm = p \pm \sqrt{p^2 - q}$. This expression is completely equivalent to Equation 10.51.

This can be solved in two steps. First solve for $s(x)$, the bracketed term in Equation 10.52,

$$(\mathcal{D}_{op} - r_+)s(x) = 0. \qquad (10.53)$$

This is a simple first-order equation with the solution

$$s(x) = Ae^{r_+x}, \qquad (10.54)$$

where A is our first undetermined constant. The second step is to solve the remainder of Equation 10.52:

$$(\mathcal{D}_{op} - r_-)y(x) = s(x) = Ae^{r_+x}. \qquad (10.55)$$

This is another first-order equation. Its solution can be obtained by using an integration factor of e^{-r_-x}, with the result

$$y(x) = e^{r_-x}\left[B + A\int dx\, e^{(r_+ - r_-)x}\right], \qquad (10.56)$$

where we have left the integral indefinite and introduced B, our second undetermined constant.

At this point, there are two cases to consider. First, if $p^2 = q$, the two "roots" are equal: $r_- = r_+ = p$. In this case, Equation 10.56 becomes

$$y(x) = (Ax + B)e^{px} \qquad \text{for } p^2 = q. \qquad (10.57)$$

On the other hand, if $p^2 \neq q$, then we get

$$y(x) = Ae^{r_+x} + Be^{r_-x} \qquad \text{for } p \neq q^2. \qquad (10.58)$$

In both cases, there are two arbitrary constants, which are determined by the boundary conditions of the problem.

"What if the roots are imaginary?" we hear you asking. In other words, what happens if $p^2 < q$, forcing r_+ and r_- to be complex? It turns out Equation 10.58 is still entirely valid, if you remember the complex definition of the exponential function:

$$e^{(x+iy)} \equiv e^x[\cos y + i \sin y]. \qquad (10.59)$$

If $p^2 < q$, r_+ and r_- will always be complex conjugates. In this case, define $r_\pm \equiv r_r \pm ir_i$ where r_r is the real part of r_+, and r_i is the imaginary part of r_+. Using this definition and redefining the arbitrary constants converts Equation 10.58 into

$$y(x) = e^{r_r x}\left[A\cos(r_i x) + B\sin(r_i x)\right] \text{ for } p^2 < q. \qquad (10.60)$$

Example 10.4 Consider the classic example of the harmonic oscillator, governed by the second-order equation

$$\frac{d^2 y(t)}{dt^2} = -\omega_o^2 y(t). \tag{10.61}$$

In operator notation, this equation is represented by

$$(\mathcal{D}_{op}^2 + \omega_o^2) y(t) = 0. \tag{10.62}$$

Referring to Equation 10.50, $p = 0$ and $q = \omega_o^2$. The roots associated with this equation are $r_{\pm} = \pm i\omega_o$. When substituted into Equation 10.60, the general solution for the simple harmonic oscillator becomes

$$y(t) = A \cos(\omega_o t) + B \sin(\omega_o t). \tag{10.63}$$

10.3.2 Finding a Second Solution Using the Wronskian

The Wronskian is a useful tool for generating the second solution of a second-order, linear, homogeneous equation when the first solution is already known. The Wronskian, itself, is a function of the independent variable and can be determined directly from the differential equation. The Wronskian for higher-order equations is discussed in a later section of this chapter

As we pointed out earlier, the most general form of a second-order, linear, ordinary, homogeneous differential equation can be written as

$$\frac{d^2 y(x)}{dx^2} + P(x)\frac{dy(x)}{dx} + Q(x) y(x) = 0. \tag{10.64}$$

Because this is a second-order, linear equation, there will be two nontrivial solutions, $y_1(x)$ and $y_2(x)$. Any linear combination of these solutions

$$y(x) = C_1 y_1(x) + C_2 y_2(x), \tag{10.65}$$

where C_1 and C_2 are arbitrary constants, is also a valid solution.

The Wronskian $W(x)$ is defined in terms of these two, nontrivial solutions as

$$W(x) = y_1(x) \left[\frac{dy_2(x)}{dx} \right] - y_2(x) \left[\frac{dy_1(x)}{dx} \right]$$
$$= y_1(x) y_2{}'(x) - y_2(x) y_1{}'(x), \tag{10.66}$$

where we have used the shorthand notation $y'(x) \equiv dy/dx$ to represent derivatives. The Wronskian can also be written using a determinant:

$$W(x) = \left| \begin{array}{cc} y_1(x) & y_2(x) \\ y_1{}'(x) & y_2{}'(x) \end{array} \right|_{det}. \tag{10.67}$$

The Wronskian can be used to tell whether the $y_1(x)$ and $y_2(x)$ solutions are independent. If two functions are linearly dependent, we can write

$$y_1(x) = k_o\, y_2(x) \tag{10.68}$$

for some choice of the constant k_o. Taking the derivative of both sides gives

$$y_1'(x) = k_o\, y_2'(x). \tag{10.69}$$

Consequently, if the two solutions are linearly dependent,

$$\frac{y_1(x)}{y_2(x)} = \frac{y_1'(x)}{y_2'(x)} \tag{10.70}$$

and $W(x) = 0$.

As a simple example, return to the harmonic oscillator equation

$$\frac{d^2 y(x)}{dx^2} + \omega_o{}^2 y(x) = 0. \tag{10.71}$$

We already have shown that the two nontrivial solutions are

$$y_1(x) = \sin(\omega_o x) \tag{10.72}$$

$$y_2(x) = \cos(\omega_o x). \tag{10.73}$$

The Wronskian in this case is

$$W(x) = \left| \begin{array}{cc} \sin(\omega_o x) & \cos(\omega_o x) \\ \omega_o \cos(\omega_o x) & -\omega_o \sin(\omega_o x) \end{array} \right|_{det}$$

$$= -\omega_o. \tag{10.74}$$

Because the Wronskian for this differential equation is not zero, the two solutions are linearly independent.

An important feature of the Wronskian is that it can be obtained directly from the differential equation, without knowing either of the solutions. Consider the general second-order differential equation given by Equation 10.64. Start with $W(x)$, as defined by Equation 10.66, and take its derivative

$$W'(x) = y_1\, y_2'' + y_1'y_2' - y_2'y_1' - y_2\, y_1''$$

$$= y_1\, y_2'' - y_2\, y_1''. \tag{10.75}$$

Both $y_1(x)$ and $y_2(x)$ are unknown functions, which separately satisfy the differential equation. The next step involves writing the pair of equations

$$y_1(x)[y_2'' + P(x)y_2' + Q(x)y_2] = 0 \tag{10.76}$$

and

$$y_2(x)[y_1'' + P(x)y_1' + Q(x)y_1] = 0, \tag{10.77}$$

which are generated by plugging the two solutions into Equation 10.64. Subtracting these two equations gives

$$y_1 y_2'' - y_2 y_1'' + P(x) \left[y_1 y_2' - y_2 y_1'\right] = 0. \tag{10.78}$$

Notice the dependence on $Q(x)$ has conveniently canceled out. Equation 10.78 can now be written in terms of the Wronskian and its derivative,

$$\frac{dW(x)}{dx} + P(x) W(x) = 0. \tag{10.79}$$

So, without any knowledge of the solutions to the original differential equation, a simple first-order equation for the Wronskian has been derived. The solution for $W(x)$ can be obtained by separation of variables:

$$\int_{W(x_o)}^{W(x)} \frac{dW}{W} = - \int_{x_o}^{x} dx\, P(x)$$

$$W(x) = W_o\, e^{- \int_{x_o}^{x} dx\, P(x)}, \tag{10.80}$$

where W_o is a constant, equal to $W(x_o)$. From this expression it is evident that if $W(x)$ is zero for some value of x, it will be zero for all x.

This ability to determine $W(x)$ directly from the differential equation is important, because we can use it to generate the second solution of a differential equation if one solution is already known. To see this, assume that $y_1(x)$ is known and $y_2(x)$ is not. This situation might arise when it is possible to guess one solution. Start by forming the derivative of $y_2(x)$ divided by $y_1(x)$:

$$\frac{d}{dx}\left[\frac{y_2(x)}{y_1(x)}\right] = \frac{y_1(x) y_2'(x) - y_2(x) y_1'(x)}{y_1^2(x)}. \tag{10.81}$$

Next express this derivative in terms of the Wronskian:

$$\frac{d}{dx}\left[\frac{y_2(x)}{y_1(x)}\right] = \frac{W(x)}{y_1^2(x)}. \tag{10.82}$$

Integrating both sides with respect to x gives

$$y_2(x) = y_1(x) \int dx\, \left[\frac{W(x)}{y_1^2(x)}\right]. \tag{10.83}$$

10.3.3   The Wronskian and Nonhomogeneous Equations

The Wronskian can also be used to obtain the particular solution to a second-order, linear, nonhomogeneous differential equation. Consider a general, linear, nonhomogeneous equation with the form:

$$a(x)\frac{d^2 y(x)}{dx^2} + b(x)\frac{dy(x)}{dx} + c(x) y(x) = d(x). \tag{10.84}$$

The functions, $a(x)$, $b(x)$, $c(x)$, and $d(x)$ are known functions of the single independent variable. The general solution to Equation 10.84 can always be written as

$$y(x) = y_p(x) + C_1 y_1(x) + C_2 y_2(x), \qquad (10.85)$$

where $y_1(x)$ and $y_2(x)$ are the two independent solutions to the corresponding second-order homogeneous equation, that is Equation 10.84 with $d(x) = 0$. The function, $y_p(x)$, is called the particular solution and is any one solution to the nonhomogeneous Equation 10.84. The constants C_1 and C_2 are determined by the boundary conditions. There must be two such conditions to determine these constants and uniquely specify the problem.

We will now show that if $a(x)$, $b(x)$, $c(x)$, $d(x)$, and the two nontrivial, homogeneous solutions $y_1(x)$ and $y_2(x)$ are known, the general solution represented by $y(x)$ in Equation 10.85 can be obtained using the Wronskian. Actually, only one of the homogeneous solutions really needs to be given at the start, because Equation 10.83 can be used to determine the second. The Wronskian can be obtained using either Equation 10.66 and the two homogeneous solutions, or from Equation 10.80 with $P(x) = b(x)/a(x)$.

It is convenient to reformat the problem. Instead of attempting to find the solution $y_p(x)$ and adding it to the two homogeneous solutions, we will try to find $y(x)$, the function given by Equation 10.85, assuming it can be written as

$$y(x) = C_1(x) y_1(x) + C_2(x) y_2(x). \qquad (10.86)$$

We will seek solutions for the two functions, $C_1(x)$ and $C_2(x)$. An important thing to notice right from the start is that $C_1(x)$ and $C_2(x)$ are not unique. This is quickly evident if you assume the $y_p(x)$ of Equation 10.85 is known. On the one hand, $C_1(x)$ could equal the constant C_1, in which case $C_2(x) = C_2 + y_p(x)/y_2(x)$. On the other hand, we could have just as easily made $C_2(x)$ the constant C_2, forcing $C_1(x) = C_1 + y_p(x)/y_1(x)$. The fact that these functions are not unique will play an important role in the discussion that follows.

Insert Equation 10.86 into the Equation 10.84. To do this, two derivatives of $y(x)$ must be calculated. Using Equation 10.86, the first derivative is given by

$$\frac{dy(x)}{dx} = C_1(x) y_1{}'(x) + C_2(x) y_2{}'(x) + C_1{}'(x) y_1(x) + C_2{}'(x) y_2(x). \qquad (10.87)$$

Because $C_1(x)$ and $C_2(x)$ are not unique, we are free to impose a relationship between them, to make them unique. This relationship will be chosen to simplify Equation 10.87. In mathematics, such a relationship is often referred to as a *gauge*. In this particular case, a useful gauge is the requirement that

$$C_1{}'(x) y_1(x) + C_2{}'(x) y_2(x) = 0. \qquad (10.88)$$

This simplifies Equation 10.87 to

$$\frac{dy(x)}{dx} = C_1(x) y_1{}'(x) + C_2(x) y_2{}'(x). \qquad (10.89)$$

The second derivative of $y(x)$ becomes

$$\frac{d^2y(x)}{dx^2} = C_1(x)\, y_1''(x) + C_2(x)\, y_2''(x) + C_1'(x)y_1'(x) + C_2'(x)y_2'(x). \quad (10.90)$$

Equations 10.86, 10.89 and 10.90 can now be substituted into Equation 10.84. Taking into account that $y_1(x)$ and $y_2(x)$ both satisfy the homogeneous equation with $d(x) = 0$, a simple first-order differential equation involving $C_1(x)$ and $C_2(x)$ is obtained:

$$a(x)\left[C_1'(x)\, y_1'(x) + C_2'(x)\, y_2'(x)\right] = d(x). \quad (10.91)$$

This equation, plus the gauge given by Equation 10.88, results in a pair of coupled differential equations that determine $C_1(x)$ and $C_2(x)$ to within two arbitrary constants:

$$C_1'(x)y_1'(x) + C_2'(x)y_2'(x) = \frac{d(x)}{a(x)} \quad (10.92)$$

$$C_1'(x)y_1(x) + C_2'(x)y_2(x) = 0. \quad (10.93)$$

To solve these equations, they must be uncoupled. Equation 10.93 can be used to solve for $C_2'(x)$:

$$C_2'(x) = -C_1'(x)\left[\frac{y_1(x)}{y_2(x)}\right]. \quad (10.94)$$

Now substitute this back into Equation 10.92 to obtain

$$C_1'(x)\left[y_1'(x)\, y_2(x) - y_1(x)\, y_2'(x)\right] = y_2(x)\frac{d(x)}{a(x)}. \quad (10.95)$$

The term in brackets is $-W(x)$, the negative of the Wronskian. Thus we get the two relations,

$$\frac{dC_1(x)}{dx} = -\frac{y_2(x)\, d(x)}{W(x)\, a(x)} \quad (10.96)$$

$$\frac{dC_2(x)}{dx} = +\frac{y_1(x)\, d(x)}{W(x)\, a(x)}. \quad (10.97)$$

Integrating these two equations gives

$$C_1(x) = -\int dx\, \left[\frac{y_2(x)\, d(x)}{W(x)\, a(x)}\right] + \text{a constant} \quad (10.98)$$

$$C_2(x) = +\int dx\, \left[\frac{y_1(x)\, d(x)}{W(x)\, a(x)}\right] + \text{another constant.} \quad (10.99)$$

These expressions can now be used to generate the solution to the nonhomogeneous equation:

$$y(x) = C_1(x)\, y_1(x) + C_2(x)\, y_2(x). \quad (10.100)$$

If the constants in Equations 10.98 and 10.99 are associated with the C_1 and C_2 of Equation 10.85, the particular solution can be written

$$y_p(x) = -y_1(x)\int dx\, \left[\frac{y_2(x)\, d(x)}{W(x)\, a(x)}\right] + y_2(x)\int dx\, \left[\frac{y_1(x)\, d(x)}{W(x)\, a(x)}\right]. \quad (10.101)$$

10.3.4 Higher-Order Wronskians

The second-order Wronskian, introduced above, can be generalized to higher-order differential equations by extending the determinant definition given by Equation 10.67. Unfortunately, these higher-order Wronskian are not very useful for generating new solutions, because the math gets very complex. They are still of value, however, for determining if solutions are linearly independent.

An nth order Wronskian can be generated from the determinant

$$W(x) = \begin{vmatrix} y_1(x) & y_2(x) & \ldots & y_n(x) \\ y_1{}'(x) & y_2{}'(x) & \cdots & y_n{}'(x) \\ y_1{}''(x) & y_2{}''(x) & \cdots & y_n{}''(x) \\ \vdots & \vdots & \vdots & \vdots \\ y_1{}^{(n)}(x) & y_2{}^{(n)}(x) & \cdots & y_n{}^{(n)}(x) \end{vmatrix}_{det} . \qquad (10.102)$$

Here the $y_i(x)$ are the n solutions of a linear n^{th} order, homogeneous differential equation. As before, if $W(x) \neq 0$, the $y_i(x)$ solutions are linearly independent.

10.4 THE METHOD OF FROBENIUS

The method of Frobenius is a series approach to solving linear, homogeneous differential equations of any order. It is usually a method of last resort, because it involves a good deal of algebra and many times leaves the solution as an infinite series that cannot be placed in closed form. However, the method can be used to obtain solutions to many equations that can be solved in no other way.

The technique is probably best described by example. Consider the second-order differential equation

$$x^2 \frac{d^2y(x)}{dx^2} + 2x\frac{dy(x)}{dx} + (x^2 - 2)y(x) = 0. \qquad (10.103)$$

This is a form of Bessel's equation, which is common in physics and engineering problems with cylindrical geometries. We will try to find a solution using a power series of the form

$$y(x) = \sum_{n=0}^{\infty} c_n x^{n+s}, \qquad (10.104)$$

where the c_n are unknown coefficients, and s is a currently unknown integer, either positive or negative, which determines the starting power of the series.

When the RHS of Equation 10.104 is inserted into Equation 10.103, we obtain

$$x^2 \sum_{n=0}^{\infty} c_n(n+s)(n+s-1)x^{n+s-2} + 2x\sum_{n=o}^{\infty} c_n(n+s)x^{n+s-1}$$

$$+ (x^2 - 2)\sum_{n=0}^{\infty} c_n x^{n+s} = 0, \qquad (10.105)$$

where we have differentiated the series for $y(x)$ term by term. This approach is valid if the resulting series converges uniformly, a condition that must be checked once s and the c_n are determined.

The terms can now be grouped based on the powers of x they contain:

$$\sum_{n=0}^{\infty} c_n \left\{ \left[(n+s)(n+s-1) + 2(n+s) - 2 \right] x^{n+s} + x^{n+s+2} \right\} = 0. \quad (10.106)$$

Equation 10.106 is of the form

$$\sum_m a_m x^m = 0. \quad (10.107)$$

The only way the LHS of this equation can be zero for all x is to have all the a_m coefficients equal to zero. To isolate the various powers of x and their coefficients, rewrite Equation 10.106 as

$$\sum_{n=0}^{\infty} c_n \left[(n+s)(n+s-1) + 2(n+s) - 2 \right] x^{n+s} + \sum_{n=2}^{\infty} c_{n-2} x^{n+s} = 0. \quad (10.108)$$

The first series in Equation 10.108 has powers of x starting with x^s, i.e., x^s, x^{s+1}, x^{s+2}, ... The second sum has powers of x starting with x^{s+2}, i.e., x^{s+2}, x^{s+3}, ... So, in order to group according to powers of x, the first two terms of the first summation in Equation 10.108 have to be written out explicitly. The remaining terms can be written as a sum from $n = 2$ to $n = \infty$:

$$c_o \left[(s)(s-1) + 2s - 2 \right] x^s + c_1 \left[(s+1)(s) + 2(s+1) - 2 \right] x^{s+1}$$

$$+ \sum_{n=2}^{\infty} \left\{ c_n \left[(s+n)(s+n-1) + 2(s+n) - 2 \right] + c_{n-2} \right\} x^{n+s}$$

$$= 0. \quad (10.109)$$

Each of the terms in Equation 10.109 must be equal to zero. The coefficients of the x^s and the x^{s+1} terms form what are called the indicial equations:

$$c_o \left[(s)(s-1) + 2s - 2 \right] = 0 \quad (10.110)$$

$$c_1 \left[(s+1)(s) + 2(s+1) - 2 \right] = 0. \quad (10.111)$$

These equations determine s. If Equation 10.104 is to generate a nontrivial solution, the c_n coefficients cannot all be zero. There are several approaches that all generate the same solution. We will start off by assuming that $c_o \neq 0$. That means in order to satisfy Equation 10.110,

$$(s)(s-1) + 2s - 2 = 0. \quad (10.112)$$

This leads to two solutions for s: $s = 1$ and $s = -2$. These two different values for s mean that Equation 10.104 really represents two solutions,

$$y_1(x) = \sum_{n=0}^{\infty} a_n x^{n+1} \qquad \text{for } s = 1 \tag{10.113}$$

$$y_2(x) = \sum_{n=0}^{\infty} b_n x^{n-2} \qquad \text{for } s = -2, \tag{10.114}$$

where the c_n have been replaced by a_n and b_n to distinguish between the two solutions. The fact that there are two nontrivial solutions is no surprise since we are dealing with a second-order differential equation. Notice that the order of the indicial equations determined the number of solutions.

Forcing $s = 1$ makes Equation 10.110 true for any nonzero value of a_o. In the same manner, when $s = -2$, Equation 10.110 holds for a nonzero value of b_o. With $s = 1$, the only way Equation 10.111 can be true is if $a_1 = 0$. Likewise, if $s = -2$, the only way Equation 10.111 works is if $b_1 = 0$. Thus the indicial equations show that there are two solutions. One starts off with $a_0 \neq 0$ and $a_1 = 0$. The second has $b_0 \neq 0$ and $b_1 = 0$. Notice we could have satisfied Equations 10.110 and 10.111 with $c_0 = 0$ and $c_1 \neq 0$. This would have led to two different values for s, i.e., $s = 0$ and $s = -3$. It is not hard to see that the solutions generated with this choice are identical to the solutions we will obtain below.

What remains now is to determine the rest of the a_n and b_n coefficients. They are obtained from the $n \geq 2$ terms of Equation 10.109. For $s = 1$ and $n \geq 2$ we obtain

$$a_n\big[(n + 1)(n) + 2(n + 1) - 2\big] + a_{n-2} = 0. \tag{10.115}$$

This forms a generating equation for the a_n:

$$a_n = -\frac{a_{n-2}}{n^2 + 3n} \qquad n \geq 2. \tag{10.116}$$

A similar argument for the b_n with $s = -2$ gives

$$b_n = -\frac{b_{n-2}}{n^2 - 3n} \qquad n \geq 2. \tag{10.117}$$

Evaluating the first five coefficients for each series gives

$$
\begin{array}{ll}
a_0 = a_0 & b_0 = b_0 \\
a_1 = 0 & b_1 = 0 \\
a_2 = -a_0/10 & b_2 = b_0/2 \\
a_3 = 0 & b_3 = 0 \\
a_4 = a_0/280 & b_4 = -b_0/8.
\end{array}
$$

So the two series solutions become

$$y_1(x) = a_0\left(x - \frac{x^3}{10} + \frac{x^5}{280} - \cdots\right) \tag{10.118}$$

and

$$y_2(x) = b_0 \left(\frac{1}{x^2} + \frac{1}{2} - \frac{x^2}{8} + \cdots \right). \tag{10.119}$$

Now that these series solutions have been obtained, we must check to see if they converge. For this example, convergence can be shown using the ratio test. For $y_1(x)$, convergence occurs if

$$\lim_{n \to \infty} \left| \frac{a_n x^{n+1}}{a_{n-2} x^{n-1}} \right| < 1. \tag{10.120}$$

From Equation 10.116, this condition is met because

$$\lim_{n \to \infty} \left| \frac{x^2}{n^2 + 3n} \right| < 1 \qquad x \neq 0, \tag{10.121}$$

for all x except $x = 0$. We must check the $x = 0$ point specifically, because at this point there is a division by zero in Equation 10.120. A quick look back at Equation 10.118 with $x = 0$ shows that there are no problems here either. So $y_1(x)$ converges for all values of x.

For $y_2(x)$, convergence occurs if

$$\lim_{n \to \infty} \left| \frac{b_n x^{n-2}}{b_{n-2} x^{n-4}} \right| < 1. \tag{10.122}$$

From Equation 10.117, we have

$$\lim_{n \to \infty} \left| \frac{x^2}{n^2 - 3n} \right| < 1 \qquad x \neq 0, \tag{10.123}$$

for all $x \neq 0$. Again, we need to return to the original series in Equation 10.119 to see what happens at $x = 0$. Because this series contains a term with a negative power of x, namely b_0/x^2, $y_2(x)$ will not converge at $x = 0$.

Plots of $y_1(x)$ and $y_2(x)$ are shown in Figure 10.1. These functions have been given special names. With $a_0 = 1/3$, $y_1(x)$ becomes the spherical Bessel function of order one and with $b_0 = -1$, $y_2(x)$ becomes the spherical Neumann function of order one.
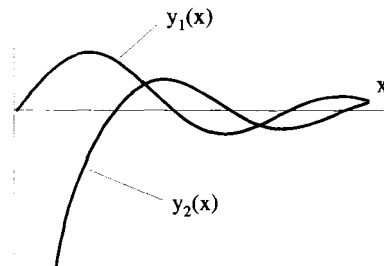

Figure 10.1   Spherical Bessel and Neumann Functions

The example given above is one of the most straightforward applications of the method of Frobenius. Complications can arise if the two indicial equations have a common value for s. This is treated in one of the exercises at the end of this chapter.

10.5   THE METHOD OF QUADRATURE

The method of quadrature sometimes works to find solutions to nonlinear, second-order differential equations of the form

$$\frac{d^2\phi(x)}{dx^2} = f(\phi), \tag{10.124}$$

where $f(\phi)$ can be a nonlinear function of the dependent variable ϕ.

10.5.1   The General Method

The general method of quadrature is fairly simple. Starting with a differential equation in the form given by Equation 10.124, both sides are multiplied by $d\phi/dx$:

$$\left[\frac{d\phi(x)}{dx}\right]\left[\frac{d^2\phi(x)}{dx^2}\right] = \left[\frac{d\phi(x)}{dx}\right]f(\phi). \tag{10.125}$$

The LHS of this equation is manipulated to give

$$\frac{1}{2}\frac{d}{dx}\left[\frac{d\phi(x)}{dx}\right]^2 = \left[\frac{d\phi(x)}{dx}\right]f(\phi). \tag{10.126}$$

Integrating both sides over x and additional manipulation gives

$$\left[\frac{d\phi}{dx}\right]^2 = 2\int d\phi\, f(\phi) \tag{10.127}$$

$$\frac{d\phi}{dx} = \left[2\int d\phi\, f(\phi)\right]^{1/2} \tag{10.128}$$

$$\int \frac{d\phi}{[2\int d\phi\, f(\phi)]^{1/2}} = \int dx \tag{10.129}$$

$$g(\phi) = x, \tag{10.130}$$

where the function $g(\phi)$ has been used to represent the result of the integration on the LHS of Equation 10.129. The solution to Equation 10.124 becomes

$$\phi(x) = g^{-1}(x). \tag{10.131}$$

To apply this method it must be possible to perform both the integrals on the RHS of Equation 10.127 and the LHS of Equation 10.129, and also to invert Equation

10.130. In this derivation, the integration constants have been ignored. As will be shown in the following example, these constants must be accounted for and make the application of this method more difficult than the steps above imply.

10.5.2 A Detailed Example: The Child–Langmuir Problem

The solution of the Child–Langmuir problem is a classic example of the use of the method of quadrature. It also demonstrates the importance of the boundary conditions for these types of solutions.

The Child–Langmuir problem describes electron flow in the anode-cathode gap of an electron gun. The geometry is shown in Figure 10.2. A heater boils electrons off the cathode, into the anode-cathode gap. In the gap, the electrons are accelerated by an electric field imposed by the applied voltage V_o. In an actual electron gun the anode is a mesh or has a hole to allow most of the accelerated electrons to continue past the anode and, in most cases, to the screen of a cathode ray tube. In this problem, however, we will concentrate only on the dynamics inside the anode-cathode gap. The cathode is located at $x = 0$, the anode at $x = d$, and the problem will be treated one dimensionally. This is a valid assumption as long as the dimensions of apparatus, perpendicular to x, are large compared to the gap distance d.

If very few electrons are in the gap at any time, their interactions with each other will be negligible. In this case, the motion of any one electron is easy to describe. It sees a uniform electric field $\overline{E}$ in the gap that arises from the linear potential $\phi(x)$, shown in Figure 10.3:

$$\overline{E} = -\overline{\nabla} \phi$$

$$= -\frac{\partial}{\partial x} \left(\frac{V_o x}{d} \right) \hat{e}_x$$

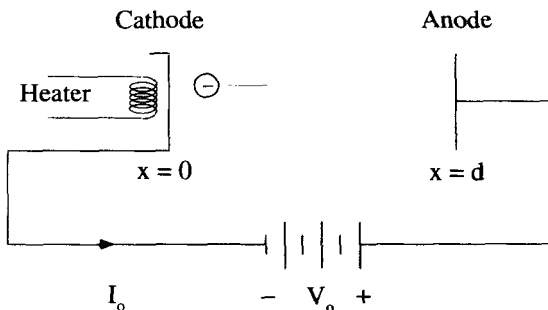$$= -\left(\frac{V_o}{d} \right) \hat{e}_x. \tag{10.132}$$

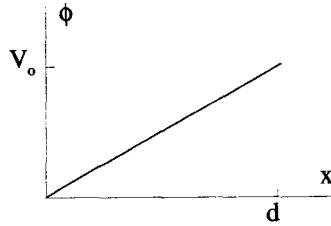**Figure 10.2** The Anode Cathode Gap of the Child–Langmuir Problem

Figure 10.3   Potential Seen by an Isolated Electron

In the gap, the electron feels a constant force $\overline{F}$ given by

$$\overline{F} = \frac{q_o V_o}{d} \, \hat{e}_x, \tag{10.133}$$

where $-q_o$ is the electron charge. The equation of motion for the electron can be written as

$$m_o \frac{d^2 x(t)}{dt^2} = \frac{q_o V_o}{d}, \tag{10.134}$$

where m_o is the electron mass. The solution to this differential equation is easily obtained by integrating twice. If it is assumed that the electron leaves $x = 0$ with zero velocity, that is $v_x(0) = 0$, its position as a function of time is given by

$$x(t) = \left(\frac{q_o V_o}{2 m_o d} \right) t^2. \tag{10.135}$$

Using conservation of energy, we can write an equation that relates the velocity of the electron, as a function of position, to the potential:

$$\frac{m_o v_x(x)^2}{2} = \int_0^x dx \, F_x$$

$$= q_o \, \phi(x). \tag{10.136}$$

To arrive at Equation 10.136 we assumed that $\phi(0) = 0$. This equation can now be solved for the velocity:

$$v_x(x) = \sqrt{\frac{2 q_o \, \phi(x)}{m_o}}. \tag{10.137}$$

In the Child–Langmuir problem, the situation is more complicated because there is not just a single electron in the anode-cathode gap. Instead, many electrons are emitted at a high, continuous rate, so the gap can be filled with billions and billions of them. The presence of the electron charges modifies the electric potential in the gap so it no longer looks as shown in Figure 10.3. The field must now be treated self-consistently using Maxwell's equations coupled with the equation of motion for the electrons.

We will look for a steady-state solution, with no time dependence. In that case, Maxwell's equations simplify to

$$\overline{\nabla} \times \overline{\mathbf{E}} = 0 \tag{10.138}$$

$$\overline{\nabla} \times \overline{\mathbf{H}} = \overline{\mathbf{J}} \tag{10.139}$$

$$\overline{\nabla} \cdot \epsilon_o \overline{\mathbf{E}} = \rho_c \tag{10.140}$$

$$\overline{\nabla} \cdot \mu_o \overline{\mathbf{H}} = 0 \tag{10.141}$$

where $\overline{\mathbf{E}}$ and $\overline{\mathbf{H}}$ are the electric and magnetic field vectors, $\overline{\mathbf{J}}$ is the current density, ρ_c is the charge density, and ϵ_o and μ_o are the dielectric constant and magnetic permiability of free space. Because there is no time dependence, the electric field has no curl and can still be expressed as the gradient of the scalar potential:

$$\overline{\mathbf{E}} = -\overline{\nabla}\phi. \tag{10.142}$$

Also, the equations involving the electric and magnetic fields uncouple, and Equations 10.140 and 10.142 can be combined, with the result

$$-\frac{d^2\phi(x)}{dx^2} = \frac{\rho_c(x)}{\epsilon_o}. \tag{10.143}$$

Given the boundary conditions of the problem, we will assume that $\phi(0) = 0$ and $\phi(d) = V_o$.

There are two unknowns in Equation 10.143: the potential $\phi(x)$ and the charge density $\rho_c(x)$, both functions of x. Consequently, to solve the problem another independent equation relating ρ_c and ϕ is necessary. This equation comes from the mechanics of the electron motion. To obtain this equation, we must introduce the current density $\overline{\mathbf{J}} = J_x(x)\hat{\mathbf{e}}_x$, which is just the charge density times the local velocity:

$$J_x(x) = \rho_c(x)\,v_x(x). \tag{10.144}$$

Using the expression derived in Equation 10.137, the velocity can be removed from Equation 10.144, and the result can be solved for $\rho_c(x)$:

$$\rho_c(x) = J_x(x)\sqrt{\frac{m_o}{2\,q_o\phi(x)}}. \tag{10.145}$$

Although we now have an equation relating the potential and the charge density, it involves another unknown $J_x(x)$, so some other information is still necessary. For that, look at the continuity equation, which we derived in Chapter 2:

$$\overline{\nabla} \cdot \overline{\mathbf{J}} + \frac{\partial \rho_c}{\partial t} = 0. \tag{10.146}$$

Since we have assumed a steady-state solution, $\partial \rho_c/\partial t = 0$, and the current density must be a constant $\overline{\mathbf{J}} = -J_o\,\hat{\mathbf{e}}_x$. We have added a minus sign to make the constant J_o positive. The actual value of J_o will be determined later, after we apply the boundary

conditions. Equation 10.145 becomes

$$\rho_c(x) = -J_o \sqrt{\frac{m_o}{2\,q_o \phi(x)}}.$$

(10.147)

Inserting Equation 10.147 into 10.143 gives:

$$\frac{d^2 \phi(x)}{dx^2} = \frac{J_o}{\epsilon_o} \sqrt{\frac{m_o}{2\,q_o \phi(x)}}.$$

(10.148)

This is a nonlinear differential equation in the form of Equation 10.124, which can now be solved using the method of quadrature. First, equation 10.148 is multiplied by $d\phi/dx$ to obtain

$$\frac{1}{2}\frac{d}{dx}\left[\frac{d\phi(x)}{dx}\right]^2 = \frac{J_o}{\epsilon_o}\sqrt{\frac{m_o}{2\,q_o}}\,\phi^{-1/2}\frac{d\phi}{dx}.$$

(10.149)

Integrating both sides from the cathode to some arbitrary point x gives

$$\left[\frac{d\phi}{dx}\right]_x^2 - \left[\frac{d\phi}{dx}\right]_{x=0}^2 = \frac{4J_o}{\epsilon_o}\sqrt{\frac{m_o}{2q_o}}\left[\phi(x)^{1/2} - \phi(0)^{1/2}\right].$$

(10.150)

Because $\phi(0)$ is zero, we can simplify this to

$$\left[\frac{d\phi}{dx}\right]_x^2 - \left[\frac{d\phi}{dx}\right]_{x=0}^2 = \frac{4J_o}{\epsilon_o}\sqrt{\frac{m_o}{2q_o}}\,\phi(x)^{1/2}.$$

(10.151)

The other constant of integration, $(d\phi/dx)^2\big|_{x=0}$, still needs to be evaluated. In order to do this, we must leave mathematics and return to the physics of the problem. According to Equation 10.142, this constant is just the square of the electric field at the cathode

$$(d\phi/dx)^2\big|_{x=0} = E_x^2(0).$$

(10.152)

If there were no electrons in the gap, the potential would behave as shown in Figure 10.3, and $(d\phi/dx)^2\big|_{x=0} = V_o^2/d^2$. With electrons in the gap, the electric field, and consequently the potential, are modified. With a single electron in the gap, the electric field of the electron adds to the vacuum electric field as shown in Figure 10.4. To the left of the electron, the total electric field is decreased slightly and to the right the field is slightly increased. As we add more electrons to the gap, regardless of how they are distributed, the electric field is always decreased at the cathode and increased at the anode, compared to the vacuum field. The modified potential has the general form shown in Figure 10.5. Of course, because electrons are point charges, the electric field changes discontinuously from one side of an electron to the other. With many electrons in the gap, however, the charge can be modeled by a continuous charge density, and the electric field and potential vary continuously.

As electrons are thermally emitted from the cathode, they are accelerated into the gap. If the electrons are emitted at a slow rate, there will be few electrons in the gap at any one time, and the electric field at the cathode will vary only slightly from the
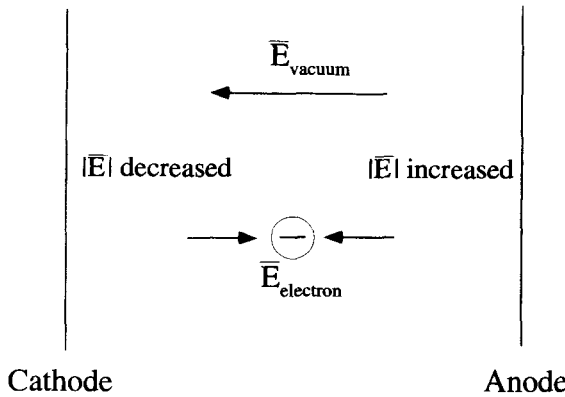
Figure 10.4 Modification of the Vacuum Electric Field by a Single Electron

vacuum case. As the temperature of the cathode is increased, electrons are emitted at a higher rate, and there will be more electrons in the gap. Now the electric field and the slope of the potential at the cathode become smaller. Therefore, there are many solutions for this problem, depending on the temperature of the cathode. The solution we will obtain is for the highest rate of electron injection possible. This will occur just as the electric field at the cathode goes to zero, because if the field were to reverse sign, electrons would be forced back into the cathode. For this solution, the second boundary condition is $(d\phi/dx)|_{x=0} = 0$.

We can now proceed with the solution. Equation 10.151 becomes

$$\frac{d\phi(x)}{dx} = \left(\frac{4J_o}{\epsilon_o}\sqrt{\frac{m_o}{2q_o}}\right)^{1/2}\phi^{1/4}(x). \tag{10.153}$$

This equation can be set up as an integral, again from the cathode where $\phi = 0$, to the point x in the gap:

$$\int_0^{\phi(x)}\frac{d\phi'}{\phi'^{1/4}} = \int_0^x dx'\left(\frac{4J_o}{\epsilon_o}\sqrt{\frac{m_o}{2q_o}}\right)^{1/2}. \tag{10.154}$$
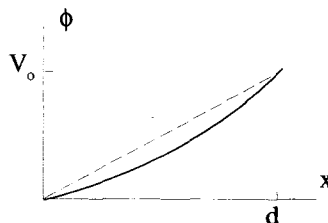

Figure 10.5 Potential Depressed by Electrons in the Gap

Performing the integration in Equation 10.154 gives the potential as a function of x:

$$\phi^{3/4}(x) = \left(\frac{9J_o}{4\epsilon_o}\sqrt{\frac{m_o}{2q_o}}\right)^{1/2} x. \qquad (10.155)$$

The RHS of this equation still has the unknown quantity J_o, but we have yet to impose the boundary condition $\phi(d) = V_o$. This condition gives

$$V_o^{3/4} = \left(\frac{9J_o}{4\epsilon_o}\sqrt{\frac{m_o}{2q_o}}\right)^{1/2} d. \qquad (10.156)$$

Solving this for J_o gives

$$J_o = \frac{4\epsilon_o}{9d^2}\sqrt{\frac{2q_o}{m_o}} V_o^{3/2} \equiv J_{cl}. \qquad (10.157)$$

The quantity J_{cl} is referred to as the Child–Langmuir current density. It depends both upon the gap distance and the applied voltage.

In this derivation, we assumed a particular cathode temperature, so that the electric field at the cathode was zero. If the cathode temperature is below this value, fewer electrons are injected into the gap, and the current density will be less than the Child-Langmuir amount given by Equation 10.157. The current density is then limited by the cathode temperature, and the gap is in what is called temperature limited flow. In this case, the current density increases exponentially with temperature until the Child-Langmuir limit is reached, as shown in Figure 10.6. This critical temperature is called the Child-Langmuir temperature T_{cl}. Above T_{cl}, electrons are emitted from the cathode at a higher rate than can be supported by the Child-Langmuir solution. If there are too many electrons in the gap, the electric field near the cathode is reversed, as shown in Figure 10.7. In this case, the field tends to decelerate the electrons, which are emitted from the cathode with a thermal distribution of velocities. The slow ones are returned to the cathode. Those that have enough velocity to make it to the point where the slope of the potential goes to zero, and the electric field reverses sign,
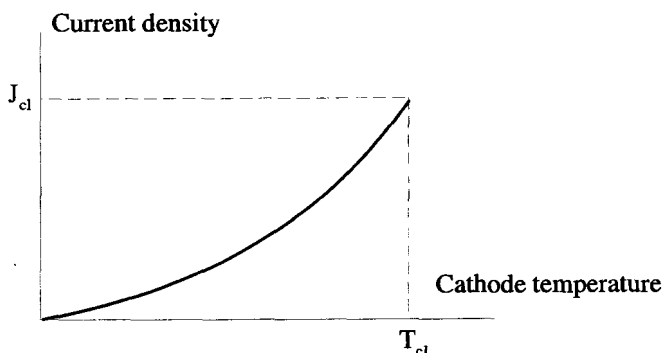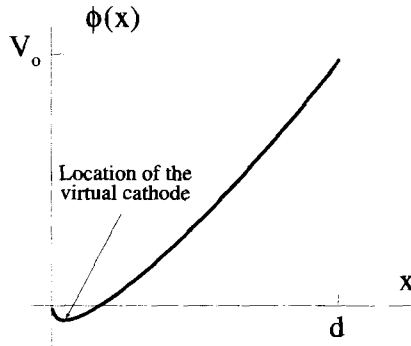

Figure 10.6 Temperature Limited Flow

Figure 10.7 Potential Distribution With the Cathode Temperature Above T_{cl}

are accelerated to the anode. The point where the slope of the potential goes to zero is called a *virtual cathode*. To the right of the virtual cathode the Child-Langmuir solution discussed above holds. In this case, however, the voltage used to determine J_{cl} from Equation 10.157 will be slightly more than the applied voltage V_o. It is increased by the magnitude of $\phi(x)$ at the virtual cathode. Also, the anode-cathode spacing used in Equation 10.157 will be less than the actual distance between the real cathode and anode. It is reduced by an amount equal to the distance from the real cathode to the virtual cathode. Consequently as the cathode temperature is raised above T_{cl}, the current density rises slightly above the J_{cl} calculated using the values V_o and d. This is a small effect, however, and the graph of the current density as a function of the cathode temperature becomes, qualitatively, as shown in Figure 10.8. Assuming the voltage and gap spacing remain constant as the temperature of the cathode is raised from zero, the current density initially increases exponentially. This is the temperature limited flow region. When the temperature reaches T_{cl}, the
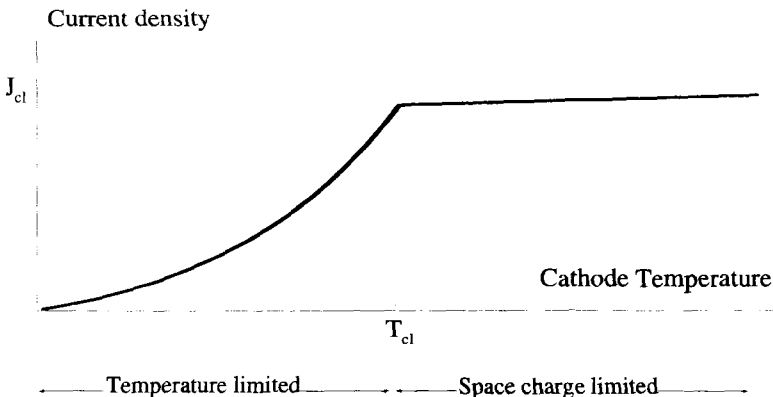

Figure 10.8 Temperature and Space Charge Limited Flow

solution of Equation 10.157 is valid. As the temperature is raised above T_{cl}, there is only a slight increase of the current density with temperature. This is called the space charge limited flow region.

10.6 FOURIER AND LAPLACE TRANSFORM SOLUTIONS

Linear differential equations with constant coefficients can many times be solved using Fourier or Laplace transform techniques. If the solutions are finite and asymptotically go to zero, they can be obtained using Fourier transforms. If they grow asymptotically, but the growth is exponentially limited, they can be handled with Laplace transforms. Solutions that grow faster than any exponential cannot be solved with these techniques.

10.6.1 Fourier Transform Solutions

We will consider the solutions to differential equations with space and time as the independent variables. If the solutions are finite and integrable, Fourier transforms with respect to the independent variables can be taken. Consider a function $f(x, t)$ that depends on both a spatial dimension x, and the time t. The spatial-time transform of this function $\underline{F}(k, \omega)$ can be obtained by a two-step process. First consider the time transform pair:

$$\underline{F}_\omega(x, \omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt\, f(x, t)e^{-i\omega t} \tag{10.158}$$

$$f(x, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega\, \underline{F}_\omega(x, \omega)e^{i\omega t}. \tag{10.159}$$

Next set up the spatial transform as

$$\underline{F}_k(k, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\, f(x, t)e^{ikx} \tag{10.160}$$

$$f(x, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\, \underline{F}_k(k, t)e^{-ikx}, \tag{10.161}$$

where we have purposely swapped the signs of the exponent from the standard Fourier transform for the reason described below. These transformations can be combined to form a pair of equations relating $f(x, t)$ and $\underline{F}(k, \omega)$:

$$\underline{F}(k, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dt\, f(x, t)e^{-i(\omega t - kx)} \tag{10.162}$$

$$f(x, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \int_{-\infty}^{\infty} d\omega\, \underline{F}(k, \omega)e^{i(\omega t - kx)}. \tag{10.163}$$

Jow you can see the purpose of the sign switch for the spatial transform. The various omponents which make up $f(x, t)$ with positive ω and positive k can be viewed s plane waves traveling in the positive x direction, with a positive phase velocity f ω/k.

Many times, in physics and engineering problems, the dependent variable is a rector quantity and there are three spatial dimensions. This complicates Equations 0.162 and 10.163 and motivates a shorthand notation for the transform pair. Consider he vector electric field $\overline{\mathbf{E}}(x, y, z, t) \equiv \overline{\mathbf{E}}(\overline{\mathbf{r}}, t)$, which is a function of three spatial limensions and time. The transform pair associated with this vector field quantity an be written as

$$\underline{\overline{\mathbf{E}}}(\overline{\mathbf{k}}, \omega) = \frac{1}{(2\pi)^2} \int_{\overline{\mathbf{r}}} d^3\overline{\mathbf{r}} \int_{-\infty}^{\infty} dt \, \overline{\mathbf{E}}(\overline{\mathbf{r}}, t) e^{-i(\omega t - \overline{\mathbf{k}} \cdot \overline{\mathbf{r}})} \tag{10.164}$$

$$\overline{\mathbf{E}}(\overline{\mathbf{r}}, t) = \frac{1}{(2\pi)^2} \int_{\overline{\mathbf{k}}} d^3\overline{\mathbf{k}} \int_{-\infty}^{\infty} d\omega \, \underline{\overline{\mathbf{E}}}(\overline{\mathbf{k}}, \omega) e^{i(\omega t - \overline{\mathbf{k}} \cdot \overline{\mathbf{r}})}. \tag{10.165}$$

n these expressions a propagation vector $\overline{\mathbf{k}}$ has been introduced. In Cartesian geom- try, it would be written as

$$\overline{\mathbf{k}} = k_x \hat{\mathbf{e}}_x + k_y \hat{\mathbf{e}}_y + k_z \hat{\mathbf{e}}_z. \tag{10.166}$$

The quantities $d^3\overline{\mathbf{r}}$ and $d^3\overline{\mathbf{k}}$ represent differential volume elements in real and $\overline{\mathbf{k}}$-space. Again, in Cartesian geometry,

$$d^3\overline{\mathbf{r}} = dx \, dy \, dz \tag{10.167}$$

$$d^3\overline{\mathbf{k}} = dk_x \, dk_y \, dk_z. \tag{10.168}$$

Notice, the integrals involving these differentials are actually three integrals spanning he entire three dimensional range of $\overline{\mathbf{r}}$ or $\overline{\mathbf{k}}$.

Example 10.5 Maxwell's equations are classic examples of differential equations nvolving space and time as the independent variables. They are used to describe the propagation of electromagnetic waves. The four equations can be written as

$$\overline{\nabla} \times \overline{\mathbf{H}}(\overline{\mathbf{r}}, t) = \frac{\partial}{\partial t} \epsilon_o \overline{\mathbf{E}}(\overline{\mathbf{r}}, t) + \overline{\mathbf{J}}(\overline{\mathbf{r}}, t)$$

$$\overline{\nabla} \times \overline{\mathbf{E}}(\overline{\mathbf{r}}, t) = -\frac{\partial}{\partial t} \mu_o \overline{\mathbf{H}}(\overline{\mathbf{r}}, t) \tag{10.169}$$

$$\overline{\nabla} \cdot \mu_o \overline{\mathbf{H}}(\overline{\mathbf{r}}, t) = 0$$

$$\overline{\nabla} \cdot \epsilon_o \overline{\mathbf{E}}(\overline{\mathbf{r}}, t) = \rho_c(\overline{\mathbf{r}}, t).$$

These equations form a set of coupled, partial differential equations with constant coefficients. If the current density $\overline{\mathbf{J}}$ and charge density ρ_c are specified, they can be looked at as driving terms, which make the equations nonhomogeneous. Alternatively,

if the current density and charge density are proportional to the fields, as is the case in linear dielectric media, the equations can be put into homogeneous form.

These equations can be transformed using the Fourier transform pair given in Equations 10.164 and 10.165. Using the fact that on transformation, $\partial/\partial t \rightarrow i\omega$ and $\partial/\partial x \rightarrow -ik_x$, Equations 10.169 become a set of coupled algebraic equations:

$$-i\overline{\mathbf{k}} \times \overline{\mathbf{H}}(\overline{\mathbf{k}}, \omega) = i\omega\epsilon_o\overline{\mathbf{E}}(\overline{\mathbf{k}}, \omega) + \overline{\mathbf{J}}(\overline{\mathbf{k}}, \omega)$$

$$-i\overline{\mathbf{k}} \times \overline{\mathbf{E}}(\overline{\mathbf{k}}, \omega) = -i\omega\mu_o\overline{\mathbf{H}}(\overline{\mathbf{k}}, \omega) \qquad\qquad (10.170)$$

$$-i\overline{\mathbf{k}} \cdot \mu_o\overline{\mathbf{H}}(\overline{\mathbf{k}}, \omega) = 0$$

$$-i\overline{\mathbf{k}} \cdot \epsilon_o\overline{\mathbf{E}}(\overline{\mathbf{k}}, \omega) = \underline{R}_c(\overline{\mathbf{k}}, \omega).$$

If the transforms $\overline{\mathbf{J}}$ and $\underline{R}_c$, are known, these equations can be solved algebraically for the transforms of the electric and magnetic fields. The real field quantities are then obtained by inverting these quantities. So you see, the task of solving the coupled partial differential equations has been simplified to simple algebra, as long and the Fourier transforms and inversions can be taken.

Example 10.6   Another problem whose solution is facilitated by the Fourier transform approach is the damped, driven, harmonic oscillator. The geometry of the problem is shown in Figure 10.9. The mass M is driven by an applied force $d(t)$. It also experiences the spring force $-K\,x(t)$ and a frictional force $-\alpha\,(dx/dt)$, proportional to its velocity. The differential equation describing the motion is

$$M\frac{d^2x(t)}{dt^2} + \alpha\frac{dx(t)}{dt} + K\,x(t) = d(t). \qquad\qquad (10.171)$$

If $d(t)$ possesses a Fourier transform $\underline{D}(\omega)$, and the solution $x(t)$ transforms into $\underline{X}(\omega)$, Equation 10.171 can be transformed to

$$\underline{X}(\omega) = \frac{-\underline{D}(\omega)/M}{\omega^2 - i\omega\alpha/M - K/M}. \qquad\qquad (10.172)$$
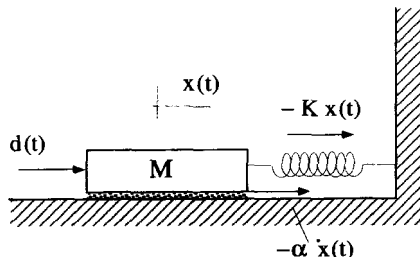

Figure 10.9   The Damped, Driven, Mass-Spring Oscillator

The denominator of the RHS of Equation 10.172 can be factored, and the solution is given by the Fourier inversion

$$x(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega \, \frac{-\underline{D}(\omega)/M}{(\omega - \underline{\omega}_+)(\omega - \underline{\omega}_-)} e^{i\omega t}, \tag{10.173}$$

where

$$\underline{\omega}_{\pm} = \frac{i\alpha}{2M} \pm \sqrt{\frac{K}{M} - \frac{\alpha^2}{4M^2}}. \tag{10.174}$$

The response $x(t)$ can be obtained once $\underline{D}(\omega)$ is determined. Let's consider the particular case where $d(t)$ is an ideal impulse, $d(t) = I_o\delta(t)$. Its transform is simply

$$\underline{D}(\omega) = I_o \frac{1}{\sqrt{2\pi}}, \tag{10.175}$$

and the response becomes

$$x(t) = \frac{1}{2\pi} \int_{\mathcal{F}} d\underline{\omega} \, \frac{-I_o/M}{(\underline{\omega} - \underline{\omega}_+)(\underline{\omega} - \underline{\omega}_-)} e^{i\underline{\omega} t}, \tag{10.176}$$

where the integral has been converted to one in the complex $\underline{\omega}$-plane on the Fourier contour $\mathcal{F}$, shown in Figure 10.10. Now we can evaluate the integral using closure and contour integration. For $t > 0$, the contour must be closed in the upper half plane, where it encloses the two poles of the integrand. For $t < 0$, it must be closed in the lower half plane and encloses no poles. A little algebra shows $x(t)$ is given by

$$x(t) = \begin{cases} 0 & t < 0 \\ \dfrac{I_o}{\sqrt{KM - \alpha^2/4}} \, e^{-\alpha t/(2M)} \sin\left(\sqrt{\dfrac{K}{M} - \dfrac{\alpha^2}{4M^2}}\, t\right) & t > 0 \end{cases}. \tag{10.177}$$
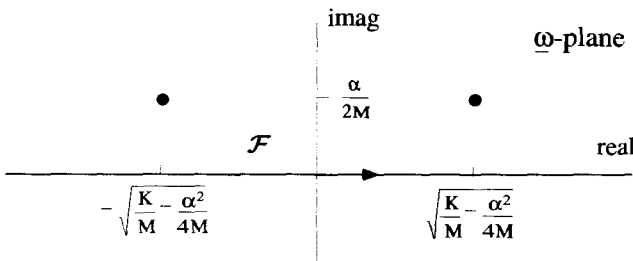

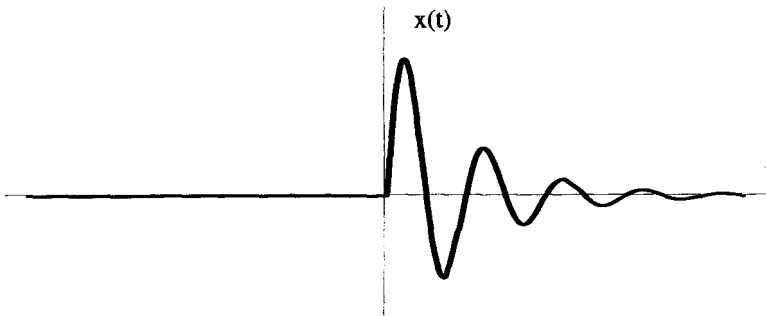**Figure 10.10** Fourier Inversion for Damped Harmonic Oscillator

Figure 10.11   Damped Harmonic Oscillator Impulse Response

This response of Equation 10.177 is plotted in Figure 10.11. The instant after the application of the impulse force at $t = 0$, the mass has acquired a finite velocity, but has not changed its position. This velocity initiates the oscillation that is damped to zero by the frictional force. Notice there is no response before $t = 0$, as you would expect, because obviously the system cannot respond until after the impulse has occurred. This kind of behavior is often referred to as being causal. Causality, a characteristic of solutions involving time, requires that there can be no response before the application of a drive.

Equation 10.171 is a second-order differential equation and needs two boundary conditions to make its solution unique. Notice that Equation 10.177 is a unique solution with no arbitrary constants. The boundary conditions must have somehow been included in the Fourier transformation and inversion process that arrived at this solution.

What are the appropriate boundary conditions for this example? For a problem of this type, the boundary conditions are typically the initial values of x and dx/dt at $t = 0$. The initial conditions in this case, however, are complicated by the δ-function located exactly at $t = 0$. There are two ways we can look at these initial conditions. First, because causality implies that $x = 0$ for $t < 0$, we can say just before the application of the impulse, at $t = 0_-$, these values are $x = 0$ and $dx/dt = 0$. Then the impulse occurs at $t = 0$ and instantaneously changes the velocity. The other view is to enforce the initial conditions just after the impulse is applied, at $t = 0_+$, and remove the impulse drive from the differential equation. In this case, these $t = 0_+$ "effective" initial conditions are $x = 0$ and $dx/dt = I_o/M$.

The solution in Equation 10.177, obtained from the Fourier analysis, satisfies both these conditions at $t = 0_-$ and $t = 0_+$. Because there are no poles of $\underline{X}(\omega)$ in the lower half ω-plane, the Fourier inversion in Equation 10.176 will always be zero for $t < 0$, and thus satisfies the $t = 0_-$ initial conditions. Looking at Equation 10.177 and its derivative for $t > 0$ shows that the $t = 0_+$ conditions are also satisfied. The instantaneous change in velocity, built into these "effective" boundary conditions, is a result of the $D(\underline{\omega})$ term in Equation 10.175.

The Fourier approach worked in this example, because the driving force could be Fourier transformed, and because the damped response could also be Fourier

ransformed. In many problems, it is not as obvious that the Fourier approach will work. For an unstable system, even if the driving term can be Fourier transformed, the response can diverge in time and, therefore, will not have a valid Fourier transform. Consequently, it is always safer to use the Laplace transform approach described below, if the possibility of instability is present.

10.6.2 Laplace Transform Solutions

The Laplace transform can also be used for solving differential equations, and is essentially the same as the Fourier approach described above. It has the advantage of being able to treat unstable systems which grow exponentially. Unfortunately, it has an added complication because the initial conditions of the problem are present in the transforms themselves and need to be treated carefully.

The standard Laplace transform pair is

$$\underline{F}(\underline{s}) = \int_0^\infty dt\, f(t) e^{-\underline{s}t} \tag{10.178}$$

$$f(t) = \frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s}\, \underline{F}(\underline{s}) e^{\underline{s}t}. \tag{10.179}$$

For the standard Laplace transform, it is assumed that $f(t)$ is zero for $t < 0$. The inversion in Equation 10.179 is performed along a Laplace contour, to the right of all the poles of $\underline{F}(\underline{s})$. This position of the Laplace contour assures that the $f(t)$ generated by the inversion is zero for $t < 0$.

As with the Fourier approach, the first step is to Laplace transform the differential equation. Let the Laplace transform operation be represented by the $\mathcal{L}$ operator:

$$\mathcal{L}\{x(t)\} = \underline{X}(\underline{s}). \tag{10.180}$$

Then from the properties of the Laplace transform,

$$\mathcal{L}\left\{ \frac{dx(t)}{dt} \right\} = \underline{s}\,\underline{X}(\underline{s}) - x(0) \tag{10.181}$$

$$\mathcal{L}\left\{ \frac{d^2 x(t)}{dt^2} \right\} = \underline{s}^2\,\underline{X}(\underline{s}) - \underline{s}\,x(0) - \left. \frac{dx(t)}{dt} \right|_{t=0}. \tag{10.182}$$

$$\vdots$$

The initial value terms on the RHS's of these equations complicate things. If, as is often the case, we know the solution and all its derivatives are zero at $t = 0$, we can drop these terms. Otherwise, they have to be carefully carried through the analysis, as discussed in the example that follows.

Example 10.7   As an example of the Laplace transform approach, consider an undamped harmonic oscillator, a simplified version of the example discussed in the previous section. The system can be described by the configuration shown in Figure 10.9, with no frictional force, i.e., $\alpha = 0$. The differential equation for this system becomes

$$M\frac{d^2x(t)}{dt^2} + K x(t) = d(t).$$
(10.183)

Applying the Laplace transformation to both sides gives

$$M\left[\underline{s}^2\underline{X}(\underline{s}) - \underline{s}x(0) - \frac{dx(t)}{dt}\bigg|_{t=0}\right] + K\underline{X}(\underline{s}) = \underline{D}(\underline{s}).$$
(10.184)

As before, let the driving force be equal to an impulse of magnitude I_o applied at $t = 0$:

$$d(t) = I_o\delta(t).$$
(10.185)

The solution should be causal, so $x(t) = 0$ for all $t < 0$. What values should be used for x and dx/dt at $t = 0$ in Equation 10.184? As with the Fourier transform example of the previous section, because of the δ-function drive at $t = 0$, there are two ways to think about these initial conditions.

The first approach defines the initial conditions at $t = 0_-$ and includes the Laplace transform of the δ-function drive in Equation 10.184. At $t = 0_-$, $x = 0$ and $dx/dt = 0$ so that Equation 10.184 becomes

$$M\underline{s}^2\underline{X}(\underline{s}) + K\underline{X}(\underline{s}) = I_o.$$
(10.186)

Now we can solve for the transformed solution:

$$\underline{X}(\underline{s}) = \frac{I_o/M}{\underline{s}^2 + K/M}.$$
(10.187)

The final step is to invert this expression,

$$x(t) = \frac{1}{2\pi i}\int_{\mathcal{L}} d\underline{s}\,\frac{I_o/M}{\underline{s}^2 + K/M}e^{\underline{s}t},$$
(10.188)

which can be evaluated using closure. The Laplace contour $\mathcal{L}$ is placed to the right of the two poles located at $\underline{s} = \pm i\sqrt{K/M}$, as shown in Figure 10.12. For $t < 0$, the contour must be closed to the right, and no poles are enclosed. When $t > 0$, however, both poles are enclosed. The residue theorem and a little algebra give the final solution:

$$x(t) = \begin{cases} 0 & t < 0 \\ \dfrac{I_o}{\sqrt{KM}}\sin(\sqrt{K/M}\,t) & t > 0 \end{cases}.$$
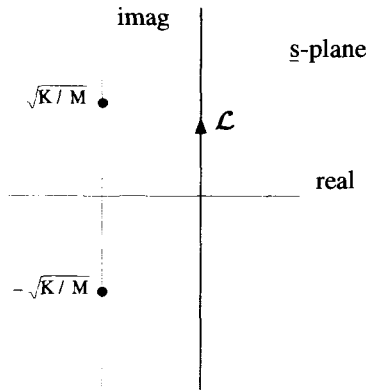(10.189)

Figure 10.12 Inversion Contour for the Harmonic Oscillator

The second approach defines the initial conditions at $t = 0_+$. In this case, the Laplace transform of the δ-function drive is not included in Equation 10.184. The initial conditions must now be evaluated just after the impulse, so that $x = 0$ and $dx/dt = I_o/M$ at $t = 0_+$. In this case, Equation 10.184 becomes

$$Ms^2\underline{X}(\underline{s}) - I_o + K\underline{X}(\underline{s}) = 0. \tag{10.190}$$

This leads to an equation for $\underline{X}(\underline{s})$ that is identical to Equation 10.187, and consequently the same solution for $x(t)$ given by Equation 10.189. Notice that if the $t = 0_+$ initial conditions as well as the Laplace transform of the δ-function drive were included in Equation 10.184, twice the correct answer for $x(t)$ would be generated.

In this example, the drive was a δ-function. Of course, this same approach works for other driving functions as well, as long as it is possible to obtain their Laplace transforms. Notice that this particular solution requires the use of the Laplace transform, as opposed to a Fourier transform. Even though the driving term can be Fourier transformed, the solution cannot. If a Fourier transform of this solution were attempted, it would have poles lying on the real $\underline{\omega}$-axis, and we could not invert it.

Example 10.8 This example explores the behavior of the unstable electrical circuit shown in Figure 10.13. This is a simple RLC-circuit except for the negative resistance, $-R_o$. This oddity is a simplified model for an active element that can actually add energy to the circuit. The circuit is driven by a sinusoidal voltage source

$$v_s(t) = \begin{cases} 0 & t < 0 \\ V_o \sin(\omega_o t) & t > 0 \end{cases}. \tag{10.191}$$

For $t < 0$ the current and all voltages in the circuit are zero.

To obtain the differential equation that describes this circuit, start by summing the voltage drops around the loop,
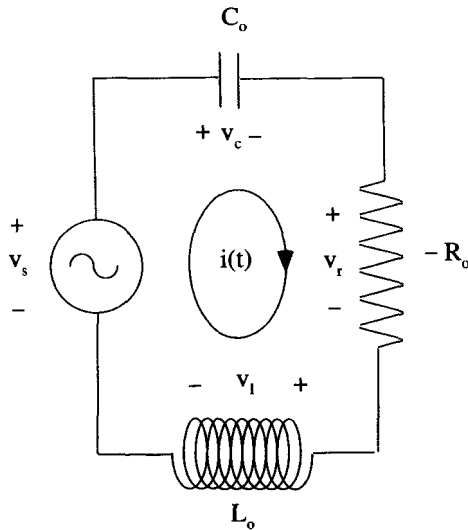
Figure 10.13   Unstable Circuit

$$v_s(t) = v_c(t) + v_r(t) + v_l(t)$$

$$= \frac{1}{C_o} \int dt\, i(t) - R_o i(t) + L_o \frac{di(t)}{dt}, \qquad (10.192)$$

where $i(t)$ is the current in the loop, as shown in Figure 10.13. Equation 10.192 is a mixed integral-differential equation. We can convert it to a pure differential equation by taking the derivative of both sides:

$$L_o \frac{d^2 i(t)}{dt^2} - R_o \frac{di(t)}{dt} + \frac{1}{C_o} i(t) = \frac{dv_s(t)}{dt}. \qquad (10.193)$$

This is a second-order, nonhomogeneous differential equation with $i(t)$ as the dependent variable. The driving term of this equation is $dv_s(t)/dt$.

Neither $v_s(t)$ nor its derivative have a valid Fourier transform, so we cannot Fourier transform Equation 10.193. We can, however, still take the Laplace transform of both sides of Equation 10.193 to get

$$\underline{s}^2 L_o \underline{I}(\underline{s}) - \underline{s} R_o \underline{I}(\underline{s}) + \frac{\underline{I}(\underline{s})}{C_o} = \underline{s}\, \underline{V}_s(\underline{s}), \qquad (10.194)$$

where $\underline{I}(\underline{s})$ and $\underline{V}_s(\underline{s})$ are the Laplace transforms of $i(t)$ and $v_s(t)$, respectively. The initial conditions for i and di/dt at $t = 0$ are zero, because causality requires the response does not occur before the application of the drive.

The Laplace transform of $v_s(t)$ is

$$\underline{V}_s(\underline{s}) = \frac{V_o\, \omega_o}{(\underline{s} + i\omega_o)(\underline{s} - i\omega_o)}, \qquad (10.195)$$

so the expression for $\underline{I}(\underline{s})$ becomes

$$\underline{I}(\underline{s}) = \frac{V_o\,\omega_o\,\underline{s}}{(\underline{s} + i\omega_o)(\underline{s} - i\omega_o)(L_o\underline{s}^2 - R_o\underline{s} + 1/C_o)}. \tag{10.196}$$

The quadratic term in the denominator can be factored as

$$L_o\underline{s}^2 - R_o\underline{s} + 1/C_o = L_o(\underline{s} - \underline{s}_+)(\underline{s} - \underline{s}_-), \tag{10.197}$$

where

$$\underline{s}_\pm = \frac{R_o}{2L_o} \pm i\sqrt{\frac{1}{L_oC_o} - \frac{R_o^2}{4L_o^2}}. \tag{10.198}$$

Notice that with a negative resistance, $\underline{s}_+$ and $\underline{s}_-$ both have a positive real part. The Laplace inversion for $i(t)$ is

$$i(t) = \frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s}\, \frac{V_o\,\omega_o\,\underline{s}\,e^{\underline{s}t}}{(\underline{s} + i\omega_o)(\underline{s} - i\omega_o)L_o(\underline{s} - \underline{s}_+)(\underline{s} - \underline{s}_-)}. \tag{10.199}$$

The Laplace contour and the poles of the integrand are shown in Figure 10.14.

To solve Equation 10.199 for arbitrary ω_o, R_o, L_o, and C_o is a messy task, which is simplified by choosing specific values for these quantities. We will let $\omega_o = 1$, $L_o = 1$, and choose the other values so that $\underline{s}_{1,2} = 1 \pm 2i$. Equation 10.199 then becomes

$$i(t) = \frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s}\, \frac{V_o\,\underline{s}\,e^{\underline{s}t}}{(\underline{s} + i)(\underline{s} - i)(\underline{s} - 1 + 2i)(\underline{s} - 1 - 2i)}. \tag{10.200}$$
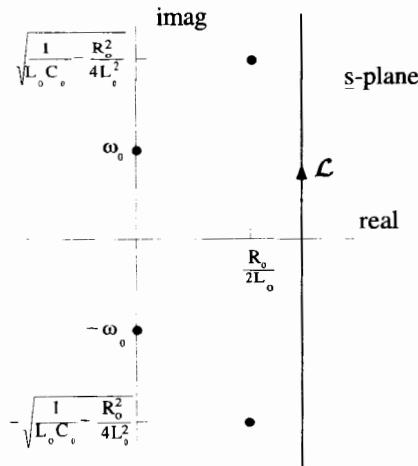

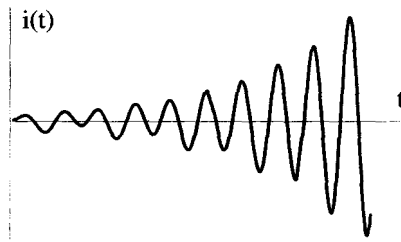**Figure 10.14** Laplace Inversion for the Unstable Circuit

Figure 10.15   Response of the Unstable Circuit

Using closure and residue techniques, the integration of Equation 10.200 can be performed to give

$$i(t) = \begin{cases} 0 & t < 0 \\ \dfrac{V_o}{5}\cos t - \dfrac{V_o}{10}\sin t - \dfrac{V_o}{5}e^t\cos(2t) + \dfrac{3V_o}{20}e^t\sin(2t) & t > 0 \end{cases}. \quad (10.201)$$

For $t > 0$ the current response can be written as

$$i(t) = \frac{V_o}{2\sqrt{5}}\cos(t + 0.15\pi) - \frac{V_o}{4}e^t\cos(2t + 0.2\pi). \quad (10.202)$$

The part of this response with a time dependence of $\cos(t + 0.15\pi)$ is a result of the driving voltage at $\omega_o = 1$. It persists at a constant amplitude for all $t > 0$. The part of the response with a time dependence of $e^t\cos(2t + 0.2\pi)$ is referred to as the characteristic response of the circuit, because it is determined entirely by R_o, C_o, and L_o. It grows exponentially in time because the negative resistance causes the circuit to be unstable. Because of this instability, the long-term behavior of the circuit is dominated by this characteristic exponential response, while the part oscillating at the drive frequency ω_o becomes negligible.

A sketch of Equation 10.202 is shown in Figure 10.15. In this figure the exponential growth rate has been decreased by an order of magnitude, i.e., we have used $e^{0.1t}$ instead of e^t, so that the effects of the two parts of the response can be seen more clearly. For short times, both components are clearly visible, but for long times the unstable part dominates.

10.7   GREEN'S FUNCTION SOLUTIONS

The Green's function approach is a powerful technique for finding solutions to linear, homogeneous, and nonhomogeneous differential equations. Two methods for obtaining Green's function solutions are discussed below. The first uses the Laplace transform and is limited to linear differential equations with constant coefficients. The second method is more general and can be applied to any linear differential equation.

10.7.1   Derivation Using the Laplace Transform

Consider an ordinary, second-order, nonhomogeneous, linear differential equation with constant coefficients

$$A_o \frac{d^2 r(t)}{dt^2} + B_o \frac{dr(t)}{dt} + C_o r(t) = d(t), \qquad (10.203)$$

where $r(t)$ is the response to a known driving function $d(t)$. The independent variable t is assumed to be time, so the concept of causality applies to the response. In other words, the response cannot occur before the application of the driving term. Assume the driving function is zero for $t < 0$, and that it has a standard Laplace transform $\underline{D}(\underline{s})$. The causal boundary conditions are $r(0) = 0$ and $dr/dt|_{t=0} = 0$. The Laplace transform of the differential equation becomes

$$(A_o \underline{s}^2 + B_o \underline{s} + C_o)\underline{R}(\underline{s}) = \underline{D}(\underline{s}), \qquad (10.204)$$

where $\underline{R}(\underline{s})$ is the Laplace transform of the response. Equation 10.204 can be solved for $\underline{R}(\underline{s})$:

$$\underline{R}(\underline{s}) = \frac{\underline{D}(\underline{s})}{A_o \underline{s}^2 + B_o \underline{s} + C_o}, \qquad (10.205)$$

which, in turn, can be inverted to give the response

$$r(t) = \frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s} \, \frac{\underline{D}(\underline{s}) e^{\underline{s}t}}{A_o \underline{s}^2 + B_o \underline{s} + C_o}. \qquad (10.206)$$

So far, this is nothing more than a recap of the Laplace transform approach to solving differential equations, which we presented in the previous section. At this point, however, we take a completely different tack.

Returning to the Laplace transform of the response, it can be looked at as the product of two functions

$$\underline{R}(\underline{s}) = \underline{D}(\underline{s})\underline{G}(\underline{s}), \qquad (10.207)$$

where $\underline{G}(\underline{s}) = 1/(A_o \underline{s}^2 + B_o \underline{s} + C_o)$. Consequently, from the product property of Laplace transforms, the time domain response is the convolution of two functions,

$$r(t) = d(t) \otimes g(t)$$

$$= \int_{-\infty}^{\infty} d\tau \, d(\tau) g(t - \tau), \qquad (10.208)$$

where $g(t)$ is the inverse Laplace transform of $\underline{G}(\underline{s})$

$$g(t) = \frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s} \, \frac{e^{\underline{s}t}}{A_o \underline{s}^2 + B_o \underline{s} + C_o}. \qquad (10.209)$$

The function $g(t)$ needs some interpretation before jumping into the convolution integral of Equation 10.208. The function $\underline{G}(\underline{s})$ satisfies the algebraic equation

$$\left(A_o \underline{s}^2 + B_o \underline{s} + C_o\right) \underline{G}(\underline{s}) = 1. \tag{10.210}$$

Therefore, the same reasoning that led to Equation 10.204 implies $g(t)$ is a solution to the differential equation

$$A_o \frac{d^2 g(t)}{dt^2} + B_o \frac{dg(t)}{dt} + C_o g(t) = \delta(t), \tag{10.211}$$

with the same boundary conditions for $g(t)$ as for $r(t)$, namely $g = 0$ and $dg/dt = 0$ at $t = 0$. The Laplace contour in Equation 10.209 is to the right of all of the poles of the integrand, so that $g(t)$ is zero for all $t < 0$. Thus, the function $g(t)$ is the response to a drive $d(t) = \delta(t)$, a unit impulse applied at $t = 0$. For this reason, $g(t)$ is sometimes called the impulse response function.

The function used in the convolution integral, Equation 10.208, is not $g(t)$ but rather $g(t - \tau)$. This is simply $g(t)$ delayed by an amount τ. Because the A_o, B_o, C_o coefficients of Equation 10.211 are not functions of time, $g(t - \tau)$ is the solution to this equation with $\delta(t)$ replaced by $\delta(t - \tau)$, and with the causal condition that $g(t - \tau)$ is zero for $t < \tau$. In other words, delaying the δ-function drive just delays the impulse response by the same amount, without changing the functional shape. This behavior is called *translational invariance*. The function $g(t - \tau)$ is called the *Green's function*. The translational invariance of this Green's function is a consequence of the constant coefficients in the differential equation and the nature of the boundary conditions. As we will show, not all Green's functions have this behavior.

The convolution in Equation 10.208 is an operation involving the driving function and the Green's function. It, in effect, sums up the contributions from all the parts of $d(t)$ to give the total response $r(t)$. Notice, once the Green's function for a specific differential equation is known, the response to *any* drive function can be determined by evaluating the integral in Equation 10.208. This is powerful stuff!

Notice this derivation required several things. First, the differential equation had to be linear, with constant coefficients. Second, the problem had to have causal boundary conditions. In the next section, we will explore a derivation which requires only that we have a linear differential equation and a particular set of *homogeneous* boundary conditions. Eventually, even this last condition will be relaxed when we will discuss Green's function solutions for problems with nonhomogeneous boundary conditions.

Example 10.9   As an example, again consider the undamped harmonic oscillator, shown in Figure 10.16, with a square pulse driving force:

$$d(t) = \begin{cases} 0 & t < 0 \\ F_o & 0 < t < T_o \\ 0 & T_o < t. \end{cases} \tag{10.212}$$
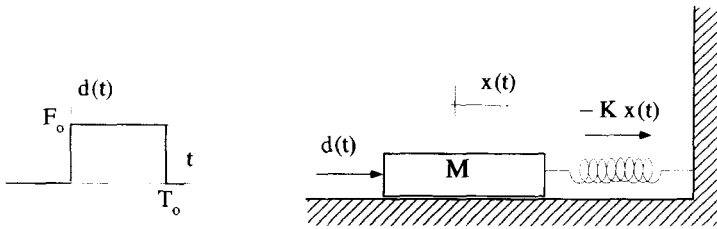
Figure 10.16   Undamped Harmonic Oscillator with Square Pulse Drive

The nonhomogeneous differential equation describing the motion is

$$M\frac{d^2x(t)}{dt^2} + K_o\, x(t) = d(t). \tag{10.213}$$

As usual, we will assume that there is no response before the drive, and impose the boundary conditions $x = 0$ and $dx/dt = 0$ at $t = 0$.

The first step in solving a problem of this type is to determine the Green's function. This is done by solving the equation

$$M\frac{d^2g(t - \tau)}{dt^2} + K_o\, g(t - \tau) = \delta(t - \tau), \tag{10.214}$$

with the same type of boundary conditions:

$$g(t - \tau) = 0 \qquad t < \tau$$

$$\frac{dg(t - \tau)}{dt} = 0 \qquad t < \tau. \tag{10.215}$$

Applying a Laplace transform to both sides of Equation 10.214 gives

$$(M\underline{s}^2 + K_o)\underline{G}(\underline{s}, \tau) = e^{-\underline{s}\tau}, \tag{10.216}$$

where $\underline{G}(\underline{s}, \tau)$ is the Laplace transform of $g(t - \tau)$. This equation can be solved for $\underline{G}(\underline{s}, \tau)$ and then Laplace inverted to give the Green's function

$$g(t - \tau) = \frac{1}{2\pi i}\int_{\mathcal{L}} d\underline{s}\frac{(1/M)e^{\underline{s}(t-\tau)}}{(\underline{s} + i\sqrt{K_o/M})(\underline{s} - i\sqrt{K_o/M})}. \tag{10.217}$$

Evaluating the integral using a Laplace contour on the right of both poles in the denominator gives

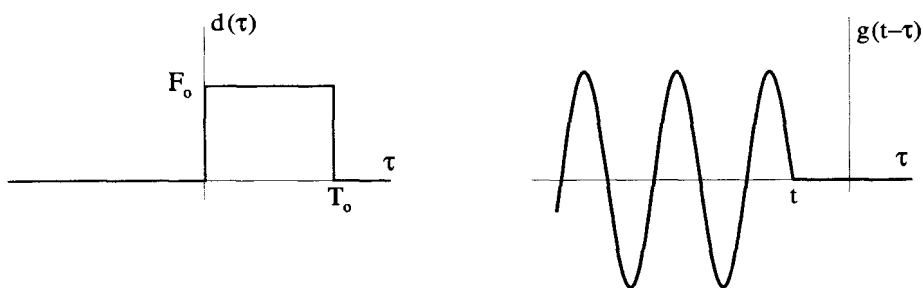$$g(t - \tau) = \begin{cases} 0 & t < \tau \\ (1/\sqrt{K_o M}) \sin\left[\sqrt{K_o/M}(t - \tau)\right] & t > \tau \end{cases}. \tag{10.218}$$

Figure 10.17 The Driving Force and Green's Function vs. τ

The next step is to use this Green's function and the driving function in Equation 10.208 to determine $x(t)$:

$$x(t) = \int_{-\infty}^{\infty} d\tau \, d(\tau)g(t - \tau). \tag{10.219}$$

To evaluate this convolution integral, it is instructive to plot $d(t)$ and $g(t - \tau)$ as functions of τ, as shown in Figure 10.17. The integral of Equation 10.219 must be evaluated for three separate ranges of t: $t < 0$, $0 < t < T_o$, and $t > T_o$. The first range is easy because, from Figure 10.17, you can see there is no overlap of $d(\tau)$ and $g(t - \tau)$, so $x(t) = 0$ for $t < 0$. In the next range, when $0 < t < T_o$, the functions overlap in the range $0 < \tau < t$, as shown in Figure 10.18. The expression for $x(t)$ becomes

$$x(t) = \int_0^t d\tau \, \frac{F_o}{\sqrt{K_oM}} \sin\left[\sqrt{K_o/M}(t - \tau)\right] \qquad 0 < t < T_o. \tag{10.220}$$

This integral is straightforward, and gives the result

$$x(t) = \frac{F_o}{K_o}\left[1 - \cos\left(\sqrt{K_o/M}t\right)\right] \qquad 0 < t < T_o. \tag{10.221}$$
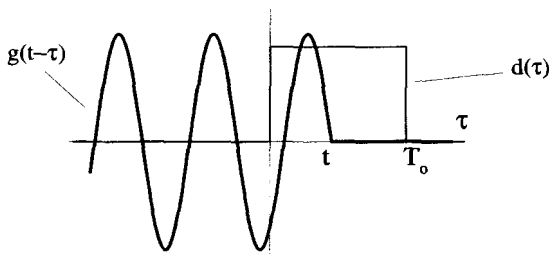
Figure 10.18 Overlap for $0 < t < T_o$
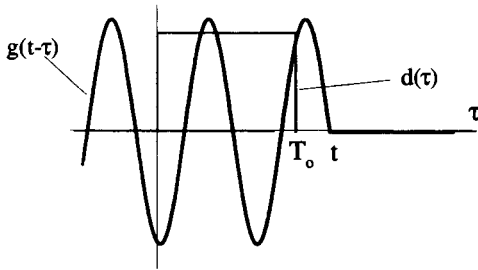
Figure 10.19   Overlap for $t > T_o$

The overlap in the final range, when $t > T_o$, is shown in Figure 10.19. In this case, the integral for $x(t)$ becomes

$$x(t) = \int_0^{T_o} d\tau \, \frac{F_o}{\sqrt{K_o M}} \sin\left[\sqrt{K_o/M}(t - \tau)\right] \qquad T_o < t < \infty, \qquad (10.222)$$

which evaluates to

$$x(t) = \frac{F_o}{K_o} \left[\sin\left(\sqrt{K_o/M}t\right) \sin\left(\sqrt{K_o/M}T_o\right)\right.$$
$$+ \cos\left(\sqrt{K_o/M}t\right) \cos\left(\sqrt{K_o/M}T_o\right)$$
$$\left. - \cos\left(\sqrt{K_o/M}t\right)\right] \qquad t > T_o. \qquad (10.223)$$

A plot of $x(t)$ is shown in Figure 10.20. Notice that $x(t)$ and $dx(t)/dt$ are continuous everywhere, and $d^2 x(t)/dt^2$ is continuous everywhere, except at $t = 0$ and $t = T_o$, where the driving force changes discontinuously.
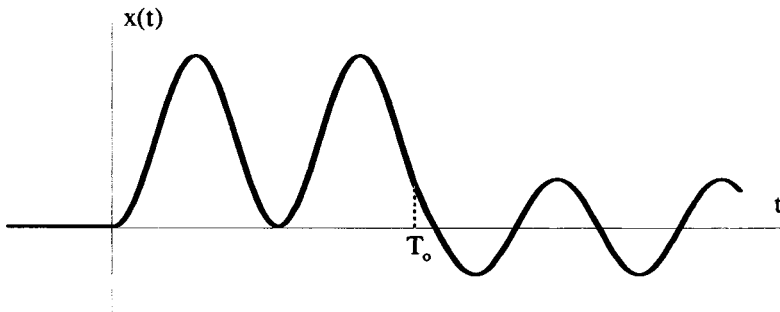
Figure 10.20   Response of Harmonic Oscillator to a Square Pulse of Force

10.7.2   A More General Approach

There is another approach to Green's functions which requires only that the differential equation be linear, and that its solutions satisfy what are called homogeneous boundary conditions. A general, second-order, nonhomogeneous equation has the form

$$A(x)\frac{d^2r(x)}{dx^2} + B(x)\frac{dr(x)}{dx} + C(x)r(x) = d(x). \qquad (10.224)$$

We are not going to assume the independent variable is causal, and to emphasize this we have used x as the independent variable instead of t. The coefficients are no longer constants, but can be a function of the independent variable. The three functions $A(x)$, $B(x)$, $C(x)$, and the driving function $d(x)$ are presumed to be known. In addition to this differential equation, two boundary conditions are necessary to make the solution $r(x)$ unique. These boundary conditions usually involve $r(x)$ or its first derivative $dr(x)/dx$ evaluated at the same or different values of x. We will begin by looking at the consequences of the linear nature of the equation and develop a Green's function solution for $r(x)$. Then the question of the boundary conditions will be addressed.

Because the differential equation we are dealing with is linear, a superposition principle applies to its solutions. That is to say, if a drive $d_1(x)$ produces the response $r_1(x)$, and another drive $d_2(x)$ produces the response $r_2(x)$, then the response to a drive function $d(x) = A_1d_1(x) + A_2d_2(x)$, where A_1 and A_2 are constants, is simply $r(x) = A_1r_1(x) + A_1r_2(x)$. This, of course, assumes the boundary conditions are satisfied for all three responses. Now assume that the drive is a δ-function $d(x) = \delta(x)$, and that this drive produces a response $r(x) = g(x)$. Next, shift the driving delta function to $x = \xi$, so that $d(x) = \delta(x - \xi)$. The response to this shifted δ-function drive will depend upon ξ, but because the coefficients of the differential equation are not constant, it cannot be assumed that the response is just $g(x)$ shifted by ξ. We will represent the response to the shifted δ-function drive by the Green's function $g(x|\xi)$, which is pronounced "g of x given ξ". This function satisfies the differential equation

$$A(x)\frac{d^2g(x|\xi)}{dx^2} + B(x)\frac{dg(x|\xi)}{dx} + C(x)g(x|\xi) = \delta(x - \xi). \qquad (10.225)$$

The superposition principle then says, if the drive is a discrete sum of weighted, shifted δ-functions, i.e.,

$$d(x) = \sum_n A_n\delta(x - \xi_n), \qquad (10.226)$$

the response is a sum of Green's functions weighted with the same A_n:

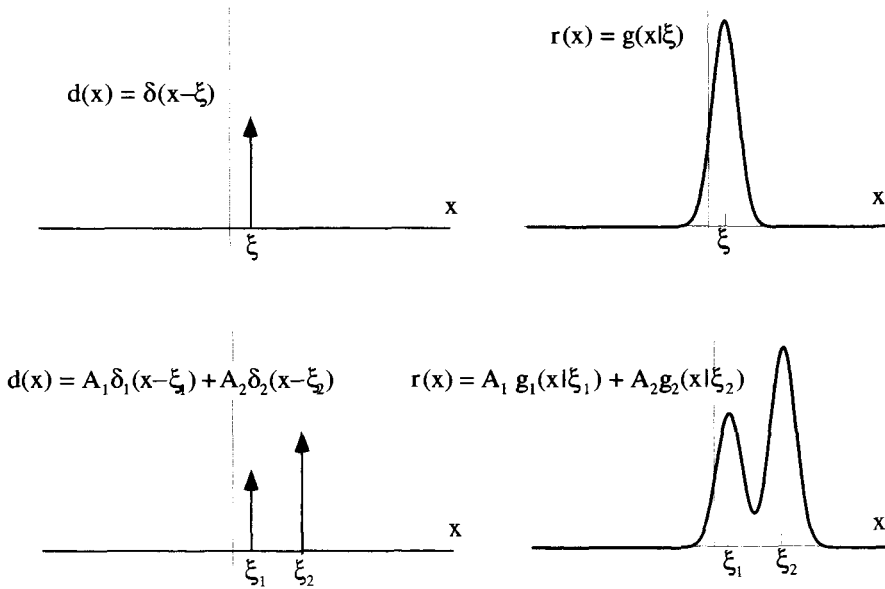$$r(x) = \sum_n A_ng(x|\xi_n). \qquad (10.227)$$

Figure 10.21   The Superposition Principal

This superposition action for a drive composed of the discrete sum of two δ-functions is shown in Figure 10.21. If $d(x)$ goes to a continuous sum of weighted δ-functions,

$$d(x) = \int_{-\infty}^{\infty} d\xi \, A(\xi)\delta(x - \xi), \qquad (10.228)$$

the response becomes a continuous sum of Green's function with the same weighting:

$$r(x) = \int_{-\infty}^{\infty} d\xi \, A(\xi)g(x|\xi). \qquad (10.229)$$

An arbitrary function $d(x)$ can always be written as a continuous sum of δ-functions:

$$d(x) = \int_{-\infty}^{\infty} d\xi \, d(\xi)\delta(x - \xi). \qquad (10.230)$$

This means that, given any driving function $d(x)$, the solution to Equation 10.224 can be written as

$$r(x) = \int_{-\infty}^{\infty} d\xi \, d(\xi)g(x|\xi). \qquad (10.231)$$

The Green's function $g(x|\xi)$ is the solution to Equation 10.225, provided the boundary conditions work out.

Equation 10.225 cannot be solved for the Green's function without specifying boundary conditions for $g(x|\xi)$. Therefore, to obtain the Green's function and evaluate the response in terms of the Green's function integral of Equation 10.231, the boundary conditions for $g(x|\xi)$ must be determined so that Equation 10.231 produces an $r(x)$ that still satisfies the original boundary conditions of the problem. Zero boundary conditions on $r(x)$ are easy to handle. By zero boundary conditions we mean that $r(x)$ or a derivative of $r(x)$ is zero at some value of x, say $x = x_o$. If this is the case, these boundary conditions on $r(x)$ will be established by a Green's function that obeys the same boundary conditions, that is $g(x|\xi)$ or its derivative is zero at the same $x = x_o$, for any ξ. These zero boundary conditions are specific examples of homogeneous boundary conditions, which are discussed below, first for the specific case of a second-order differential equations, and then in the general case.

For a second-order linear differential equation, two boundary conditions must be specified. They generally involve the response and/or its first derivative, evaluated at the same or two different values of the independent variable. The most general forms for homogeneous boundary conditions for a second-order equation are

$$\alpha_1 \, r(x)\Big|_{x=a} + \alpha_2 \frac{dr(x)}{dx}\Big|_{x=a} = 0 \qquad (10.232)$$

$$\beta_1 \, r(x)\Big|_{x=b} + \beta_2 \frac{dr(x)}{dx}\Big|_{x=b} = 0, \qquad (10.233)$$

where α_1, α_2, β_1 and β_2 are constants. If these types of boundary conditions are applied to a $g(x|\xi)$, it is easy to show that the $r(x)$ generated by Equation 10.227, and consequently Equation 10.231, will satisfy the same homogeneous boundary conditions. The homogeneous boundary conditions of the Laplace Example 10.9 fit this form with $a = b = 0$, $\alpha_1 = \beta_2 = 1$, and $\alpha_2 = \beta_1 = 0$.

For a linear, nonhomogeneous differential equation of nth order, there must be n boundary conditions that may involve the response up to its $n - 1$ derivative. For an nth-order equation, the general form of the homogeneous boundary conditions are n different equations:

$$\sum_{i=1}^{n} \alpha_i \frac{d^{i-1}r(x)}{dx^{i-1}}\Big|_{x=a} = 0. \qquad (10.234)$$

The first equation involves a set of constants α_i and the function and/or its derivatives are evaluated at $x = a$; the second, a set of constants β_i and terms evaluated at $x = b$; etc.

The Green's function argument is now complete. Given a linear, nonhomogeneous differential equation in the form of Equation 10.224 and a complete set of homogeneous boundary conditions in the form of Equation 10.234, the response can be obtained from a Green's function integral. The Green's function $g(x|\xi)$ is the solution to the original differential equation with a δ-function drive and satisfies the same set of homogeneous boundary conditions.

This derivation for the Green's function solution is correct, but a little sloppy. In Appendix D there is a more formal, but far less intuitive proof for the existence of Green's functions and for the properties of homogeneous boundary conditions. These proofs are for second-order differential equation, but are easily extended to linear equations of any order.

Example 10.10 We will now investigate a classic problem which uses a Green's function solution. In this example, space is the independent variable instead of time. Imagine a uniform, elastic string, stretched between two fixed points, as shown in Figure 10.22. The ends of the string, at $x = 0$ and $x = L$, are both held at $y = 0$. A distributed force per unit length $F(x)$ is applied along the string. Our sign convention is that a positive force points in the negative y direction. Assume the string has a uniform tension T_o, and that the displacement caused by $F(x)$ is so small that this tension does not change.

The differential equation describing the displacement of the string is

$$T_o \frac{d^2 y(x)}{dx^2} = F(x), \tag{10.235}$$

and the boundary conditions are homogeneous, with $y(0) = y(L) = 0$. Because this is a linear differential equation, it can be solved using a Green's function. The displacement becomes

$$y(x) = \int_0^L d\xi \, F(\xi) g(x|\xi), \tag{10.236}$$

where $g(x|\xi)$ is a solution to

$$T_o \frac{d^2 g(x|\xi)}{dx^2} = \delta(x - \xi), \tag{10.237}$$

with $g(0|\xi) = g(L|\xi) = 0$. Because the Green's function obeys homogeneous boundary conditions, the solution it generates will obey the same homogeneous boundary conditions.
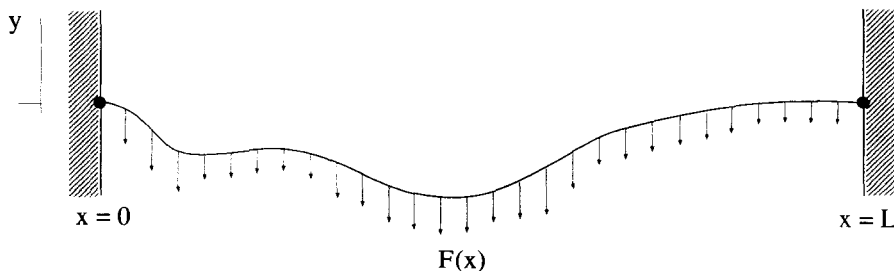

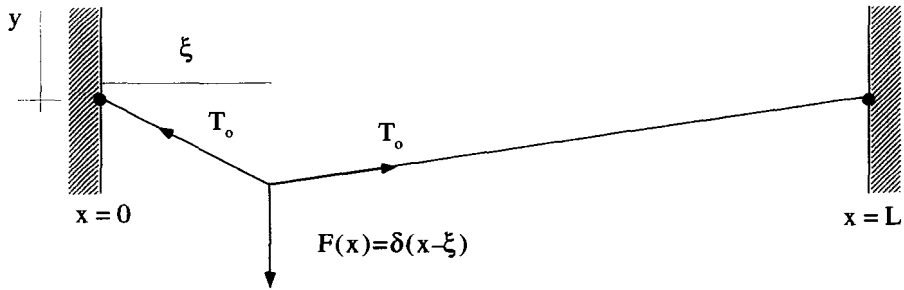**Figure 10.22** The Stretched String Problem

Figure 10.23   The Green's Function Displacement of the String

The interpretation of Equation 10.237 is simple and is depicted in Figure 10.23. The driving force is a single δ-function located at $x = \xi$. The response to that driving term is the Green's function $g(x|\xi)$, which is zero at $x = 0$ and $x = L$ for all $0 < \xi < L$. Notice, in the small displacement approximation, the vertical components of the tension add to cancel the applied δ-function force per unit length. The fact that the horizontal components of the tension do not exactly cancel is ignored.

The Green's function for this problem can be obtained by solving Equation 10.237. Everywhere, except exactly at the point $x = \xi$, the equation for $g(x|\xi)$ is

$$T_o \frac{d^2 g(x|\xi)}{dx^2} = 0, \tag{10.238}$$

which has the simple solution

$$g(x|\xi) = Ax + B. \tag{10.239}$$

Because there are two different regions that have this solution, separated by the point $x = \xi$, we can write the general solution for $g(x|\xi)$ as

$$g(x|\xi) = \begin{cases} A_1 x + B_1 & 0 < x < \xi \\ A_2 x + B_2 & \xi < x < L \end{cases}, \tag{10.240}$$

where the constants A_1 through B_2 must be determined.

The boundary conditions at $x = 0$ and $x = L$ give two equations involving these constants,

$$A_1(0) + B_1 = 0 \tag{10.241}$$

and

$$A_2(L) + B_2 = 0. \tag{10.242}$$

A third equation is obtained by requiring $g(x|\xi)$ to be continuous at $x = \xi$:

$$A_1(\xi) + B_1 = A_2(\xi) + B_2. \tag{10.243}$$

In other words, the string does not break due to the force. The fourth and final equation takes a bit more work, and involves the driving term $\delta(x - \xi)$, which has not yet been used. Start by integrating Equation 10.237 from $x = \xi - \epsilon$ to $x = \xi + \epsilon$ in the limit of $\epsilon \to 0$:

$$\int_{\xi-\epsilon}^{\xi+\epsilon} dx\, T_o \frac{d^2 g(x|\xi)}{dx^2} = \int_{\xi-\epsilon}^{\xi+\epsilon} dx\, \delta(x - \xi). \qquad (10.244)$$

The LHS of this equation goes to T_o times the discontinuity in slope of $g(x|\xi)$ at $x = \xi$, and the RHS is one because we are integrating over a δ-function. The result is

$$T_o \left(\frac{dg(x|\xi)}{dx} \bigg|_{\xi+\epsilon} - \frac{dg(x|\xi)}{dx} \bigg|_{\xi-\epsilon} \right) = 1. \qquad (10.245)$$

Taking the limit as $\epsilon \to 0$, and referring back to Equation 10.240, gives

$$T_o (A_2 - A_1) = 1. \qquad (10.246)$$

Notice that this equation is essentially a statement that the vertical forces at $x = \xi$ must cancel. Equations 10.241–10.243, along with Equation 10.246, constitute a set of four independent equations that uniquely determine A_1, A_2, B_1, and B_2. Equation 10.240 for the Green's function becomes

$$g(x|\xi) = \begin{cases} -[x/(T_o L)](L - \xi) \equiv g_1(x|\xi) & 0 < x < \xi \\ -[\xi/(T_o L)](L - x) \equiv g_2(x|\xi) & \xi < x < L \end{cases}. \qquad (10.247)$$

Notice that this Green's function does not have the simple form $g(x - \xi)$ of a translationally invariant solution. A quick look at Figure 10.23 and the boundary conditions at the ends of the string show why this must be the case.

This Green's function can now be used in Equation 10.236 to obtain the displacement of the string for an arbitrary loading force density $F(x)$. Some care, however, must be taken in setting up this integral. For this integration, x is held fixed, somewhere between 0 and L, while the integration variable ξ ranges from 0 to L. According to Equation 10.247, $g(x|\xi) = g_2(x|\xi)$ when $\xi < x$ and $g(x|\xi) = g_1(x|\xi)$ when $\xi > x$. The integration of Equation 10.236 must therefore be broken up into two parts. The Green's function solution for an arbitrary $F(x)$ becomes

$$y(x) = \int_0^x d\xi\, \frac{-\xi}{T_o L}(L - x)F(\xi) + \int_x^L d\xi\, \frac{-x}{T_o L}(L - \xi)F(\xi). \qquad (10.248)$$

10.7.3 Nonhomogeneous Boundary Conditions

Up to now, the Green's function analysis has been for problems with homogeneous boundary conditions. In this section, we will review, briefly, why homogeneous
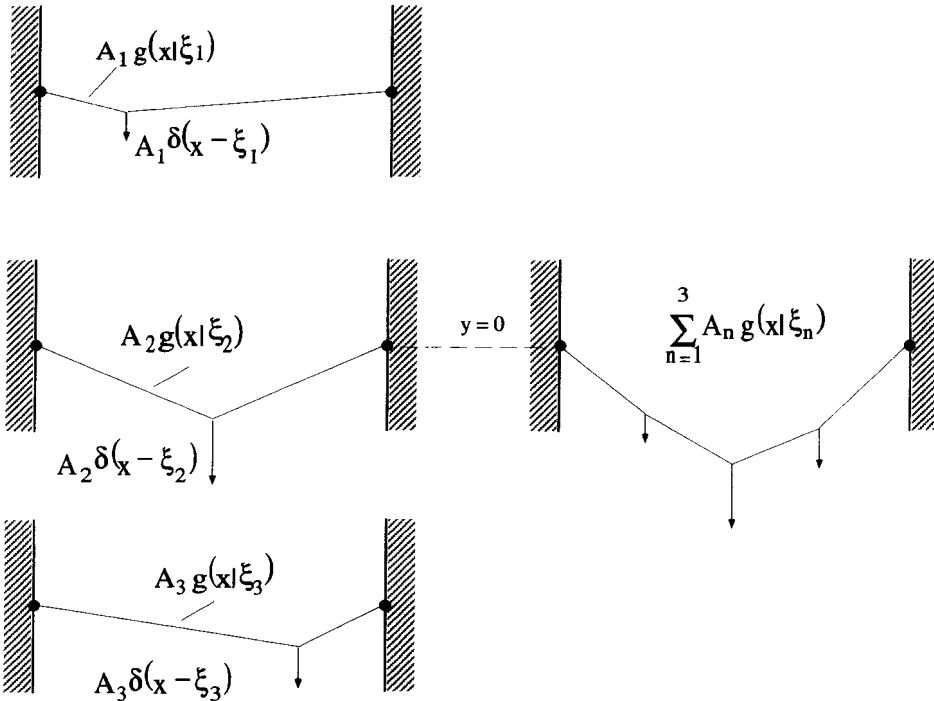
Figure 10.24   Homogeneous Boundary Conditions

boundary conditions are important. Then a method for dealing with nonhomogeneous boundary conditions will be introduced.

In the case of the stretched string problem, the example of the previous section, the homogeneous boundary conditions specified that the response be zero at either end of the string. These boundary conditions fit the general homogeneous form of Equation 10.234. We required the Green's function to obey the same boundary conditions. This works because when the Green's function is constrained to be zero at $x = 0$ and $x = L$, as shown in Figure 10.23, an arbitrary sum of Green's functions will automatically obey the same conditions. As an example of this, the sum of three Green's functions is depicted in Figure 10.24. Therefore, forcing the Green's function to obey the same boundary conditions as the original differential equation is the proper approach in this case.

The difficulty with nonhomogeneous boundary conditions is easily demonstrated with the string problem by changing the second boundary condition so that $y(L) = y_o$. A generic load and displacement which obeys this new condition is shown in Figure 10.25. Problems immediately develop if these boundary conditions are applied to the Green's function, and the response is generated with a summation of several of them. A sum of three of these Green's functions is shown in in Figure 10.26. As with Figure 10.24, here $F(x)$ is composed of three weighted δ-functions. The response to each
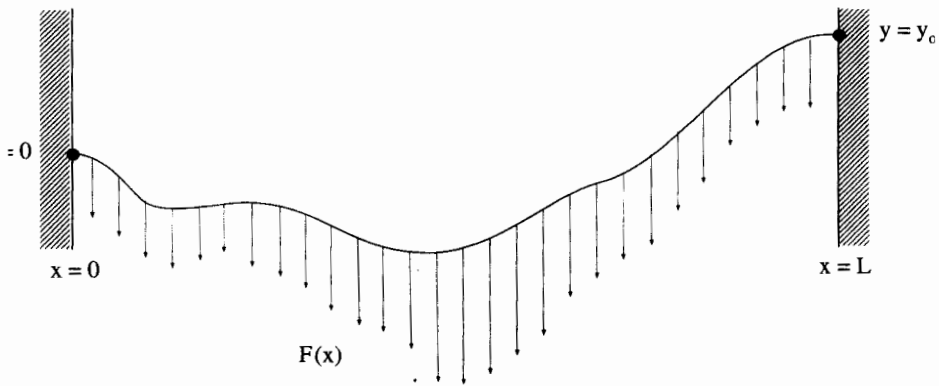
Figure 10.25   Loaded String with Nonhomogeneous Boundary Conditions

of these individual forces is shown on the left, each one proportional to a Green's function which satisfies the nonhomogeneous boundary conditions. If these solutions are simply summed, as would be done by the Green's function integral, the result would not satisfy the required boundary conditions. This is shown on the right of Figure 10.26 where the three terms have been summed to give $y(L) = 3y_o$. Clearly, if
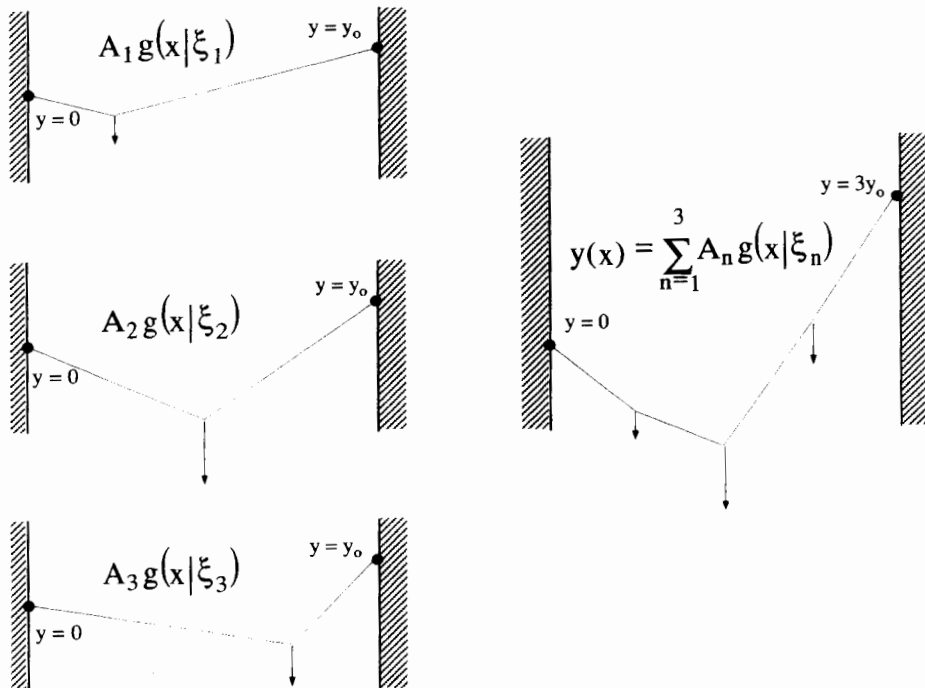

Figure 10.26   The Nonhomogeneous Boundary Condition Problem

$g(x|\xi)$ satisfies the nonhomogeneous boundary conditions, the correct answer is not generated by the Green's function sum.

Fortunately, there is a way around this problem. It is clear from the picture above that to keep the generic Green's function integral

$$r(x) = \int_a^b d\xi \, d(\xi) g(x|\xi) \qquad (10.249)$$

from diverging for $x = a$ or $x = b$, the Green's function must be forced to have homogeneous boundary conditions. But we can add to this function another term, which makes the solution obey the nonhomogeneous boundary conditions.

To see how this works, let's return to the string problem with the nonhomogeneous boundary conditions $y(0) = 0$ and $y(L) = y_o$. The solution $y(x)$ still satisfies the differential equation

$$T_o \frac{d^2 y(x)}{dx^2} = F(x). \qquad (10.250)$$

Now break $y(x)$ into two parts,

$$y(x) = y_1(x) + y_2(x), \qquad (10.251)$$

where $y_1(x)$ satisfies the nonhomogeneous differential equation, Equation 10.250, but has homogeneous boundary conditions, i.e., $y_1(0) = 0$ and $y_1(L) = 0$. Let $y_2(x)$ satisfy the homogeneous version of Equation 10.250

$$T_o \frac{d^2 y_2(x)}{dx^2} = 0, \qquad (10.252)$$

with the nonhomogeneous boundary conditions, $y_2(0) = 0$ and $y_2(L) = y_o$. The sum of these two solutions, $y_1(x) + y_2(x)$, will satisfy *both* the nonhomogeneous differential equation *and* the nonhomogeneous boundary conditions. The $y_1(x)$ part of the solution can be found with the standard Green's function approach, by requiring the Green's function to satisfy homogeneous boundary conditions. The $y_2(x)$ part is obtained by solving a simple homogeneous differential equation and then applying the nonhomogeneous boundary conditions.

In this case, $y_1(x)$ is the same solution found in Equation 10.248:

$$y_1(x) = \int_0^x d\xi \, F(\xi) \frac{-\xi}{T_o L}(L - x) + \int_x^L d\xi \, F(\xi) \frac{-x}{T_o L}(L - \xi). \qquad (10.253)$$

The $y_2(x)$ part of the solution, the solution to Equation 10.252, is simply

$$y_2(x) = Ax + B. \qquad (10.254)$$

Applying the nonhomogeneous boundary conditions forces the constants to be $A = y_o/L$ and $B = 0$, so that the total solution is

$$y(x) = \int_0^x d\xi \, F(\xi) \frac{-\xi}{T_o L}(L - x) + \int_x^L d\xi \, F(\xi) \frac{-x}{T_o L}(L - \xi) + \frac{y_o}{L} x. \qquad (10.255)$$

In summary, to solve a linear, nonhomogeneous differential equation with non-homogeneous boundary conditions, break the problem into two parts. Use a Green's function approach to solve the original, nonhomogeneous differential equation using a Green's function which satisfies homogeneous boundary conditions. The desired solution is then the sum of this solution and the solution to a homogeneous version of the differential equation, which obeys the nonhomogeneous boundary conditions.

10.7.4 Multiple Independent Variables

To this point, all our Green's function examples have had a single independent variable. How are things modified when we consider higher dimensions? In this section, through examples, we will explore two-dimensional solutions for two common problems, diffusion and wave propagation. Both examples have one space dimension and the time dimension.

Example 10.11 The Green's function approach can also be used to solve linear partial differential equations. Consider the diffusion of heat along the infinite, one-dimensional conducting rod shown in Figure 10.27. The homogeneous diffusion equation

$$D^2 \frac{\partial^2 T(x,t)}{\partial x^2} - \frac{\partial T(x,t)}{\partial t} = 0 \qquad (10.256)$$

describes the time and space dependence of the temperature $T(x,t)$ along the rod. The constant D is called the diffusion coefficient. A common way to formulate a diffusion problem is to specify $T(x,0) \equiv T_o(x)$, the initial temperature distribution at $t = 0$, as shown in Figure 10.27, and then solve for $T(x,t)$ for $t > 0$. It will be assumed that $T(x,t) \to 0$ as $x \to \pm\infty$ for all t.
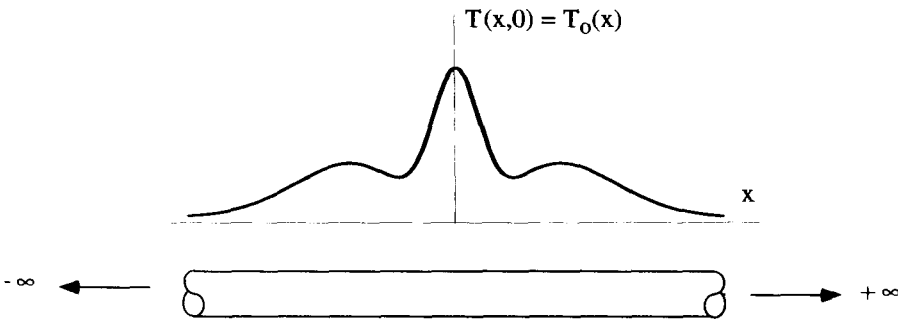

Figure 10.27 One-Dimensional Conducting Rod with Initial Temperature Distribution

At first, this does not look like a problem that could be solved using a Green's function method, since the differential equation is homogeneous and there is no driving term. It is, however, a linear equation, so the superposition principle does apply. That is to say, if $T_o(x) = \delta(x - \xi)$ results in $T(x, t) = g(x|\xi, t)$, then

$$T_o(x) = \sum_n A_n \delta(x - \xi_n) \tag{10.257}$$

will result in

$$T(x, t) = \sum_n A_n g(x|\xi_n, t) \tag{10.258}$$

for all $t > 0$. Taking these summations to continuous integrals allows us to argue that, if

$$T_o(x) = \int_{-\infty}^{\infty} d\xi \, T_o(\xi) \delta(x - \xi), \tag{10.259}$$

then

$$T(x, t) = \int_{-\infty}^{\infty} d\xi \, T_o(\xi) g(x|\xi, t) \tag{10.260}$$

for all $t > 0$. Notice how we have set up the arguments of the Green's function. The variables x and ξ have been used as standard Green's function variables, while t is not involved in the Green's function integral of Equation 10.260. This is indicated by the notation in the argument of the Green's function. Only variables immediately to the right of a "|" will take part in Green's function integrations.

So the solution to the problem can be formulated using a Green's function, with the initial temperature distribution acting as the weighting function in the Green's function integral. The boundary conditions require $T(x, t) \to 0$ as $x \to \pm\infty$ and are homogeneous. Consequently, the same boundary conditions are applied to $g(x|\xi, t)$. This Green's function must satisfy

$$D^2 \frac{\partial^2 g(x|\xi, t)}{\partial x^2} - \frac{\partial g(x|\xi, t)}{\partial t} = 0, \tag{10.261}$$

with the initial condition at $t = 0$

$$g(x|\xi, 0) = \delta(x - \xi) \tag{10.262}$$

and the boundary conditions described above. Notice in this problem, the initial condition is treated very differently from the boundary conditions.

One method to solve Equation 10.261 is to use a Fourier transform in space. A standard Laplace transform cannot be used because the solution must be valid for all x. We define

$$\underline{G}(k|\xi, t) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \, g(x|\xi, t) e^{-ikx}, \tag{10.263}$$

and the Fourier transform of Equation 10.261 becomes

$$- D^2 k^2 \underline{G}(k|\xi, t) = \frac{\partial \underline{G}(k|\xi, t)}{\partial t}. \tag{10.264}$$

This is an easy first-order equation, with the general solution

$$\underline{G}(k|\xi, t) = \underline{G}_o\, e^{-D^2 k^2 t}, \tag{10.265}$$

where the complex constant $\underline{G}_o \equiv \underline{G}(k|\xi, 0)$. But $g(x|\xi, 0) = \delta(x - \xi)$, so

$$\begin{aligned}
\underline{G}_o &= \mathcal{F}\{g(x|\xi, 0)\} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\, \delta(x - \xi) e^{-ikx} \\
&= \frac{1}{\sqrt{2\pi}} e^{-ik\xi}.
\end{aligned} \tag{10.266}$$

The Fourier transform of the Green's function becomes

$$\underline{G}(k|\xi, t) = \frac{1}{\sqrt{2\pi}} e^{-ik\xi} e^{-D^2 k^2 t}, \tag{10.267}$$

and the Green's function, itself, is the Fourier inversion of this expression:

$$g(x|\xi, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\, \frac{1}{\sqrt{2\pi}} e^{-ik\xi} e^{-D^2 k^2 t} e^{ikx}. \tag{10.268}$$

This inversion can be done most easily using a convolution trick. The Fourier transform of this Green's function can be looked at as the product of two functions

$$\underline{G}(k|\xi, t) = \underline{F}(k|\xi, t)\underline{H}(k|\xi, t), \tag{10.269}$$

where

$$\underline{F}(k|\xi, t) = \frac{1}{\sqrt{2\pi}} e^{-ik\xi} \tag{10.270}$$

and

$$\underline{H}(k|\xi, t) = e^{-D^2 k^2 t}. \tag{10.271}$$

The Green's function is therefore $1/\sqrt{2\pi}$ times the convolution of the Fourier inversions of Equations 10.270 and 10.271. Convolution is usually a messy process, but not in this case because the inversion of 10.270 is just a δ-function

$$\mathcal{F}^{-1}\left\{ \frac{1}{\sqrt{2\pi}} e^{-ik\xi} \right\} = \delta(x - \xi), \tag{10.272}$$

and δ-functions convolve easily. The second function, Equation 10.271, is a Gaussian and it inverts to another Gaussian:

$$\mathcal{F}^{-1}\{e^{-D^2k^2t}\} = \frac{1}{\sqrt{2D^2t}}e^{-x^2/(4D^2t)}. \tag{10.273}$$

The Green's function becomes a convolution over x:

$$
\begin{aligned}
g(x|\xi, t) &= \frac{1}{\sqrt{2\pi}} \left[\frac{1}{\sqrt{2D^2t}}e^{-x^2/(4D^2t)} \otimes \delta(x - \xi) \right] \\
&= \frac{1}{\sqrt{4\pi D^2 t}} \int_{-\infty}^{\infty} dx'\, e^{-(x-x')^2/(4D^2t)}\, \delta(x' - \xi) \\
&= \frac{1}{\sqrt{4\pi D^2 t}}\, e^{-(x-\xi)^2/(4D^2t)}. \tag{10.274}
\end{aligned}
$$

Equation 10.274 describes how an initial temperature distribution $T(x, 0) = \delta(x - \xi)$ evolves in time and space. A graph of this function is shown in Figure 10.28. Notice that in this case $g(x|\xi, t)$ has the form $g(x - \xi, t)$, and thus the Green's function is translationally invariant. As the initial δ-function shifts in space, the response shifts correspondingly. This happens because the differential equation has constant coefficients and the homogeneous boundary conditions are located at infinity.
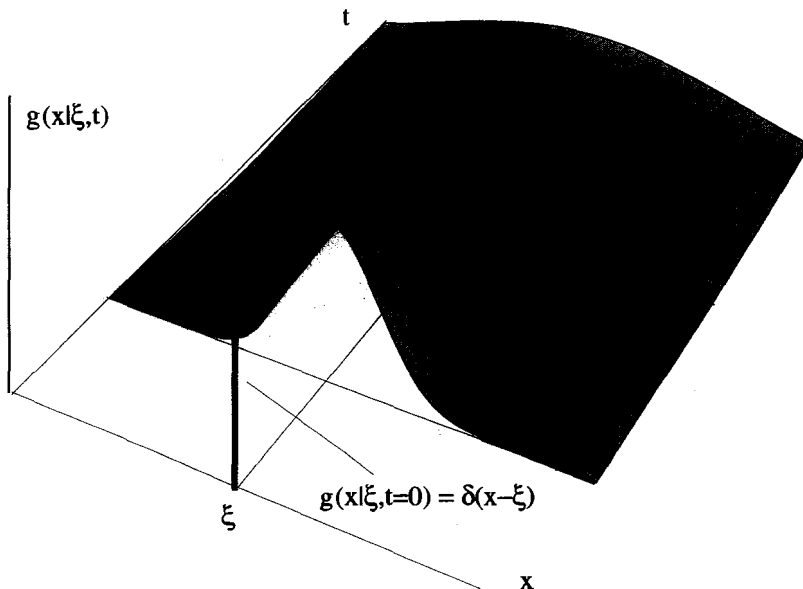

Figure 10.28   Green's Function for the Diffusion Equation

The general response for an arbitrary initial temperature distribution $T_o(x)$ is then

$$T(x, t) = \frac{1}{\sqrt{4\pi D^2 t}} \int_{-\infty}^{\infty} d\xi \, T_o(\xi) \, e^{-(x-\xi)^2/(4D^2 t)}. \tag{10.275}$$

Notice the integration of Equation 10.275 is over only the x variable. The t variable just hangs around as a simple time dependence of the Green's function. In the next example, both time and space are treated as Green's function variables, and the response will involve an integral over both of these quantities.

Example 10.12 We now consider the same one-dimensional, conducting rod as we used in the previous example, but this time with an external driving term. Imagine a special heating element, running through the center of the rod, which can add heat in an arbitrary way that depends both on x and t. The heat source $s(x, t)$ turns Equation 10.256 into a nonhomogeneous differential equation:

$$\frac{\partial T(x, t)}{\partial t} - D^2 \frac{\partial^2 T(x, t)}{\partial x^2} = s(x, t). \tag{10.276}$$

The signs have been adjusted in this equation so that a positive source term produces a positive change in temperature. The solution still must obey the homogeneous boundary conditions, $T(x, t) \rightarrow 0$ as $x \rightarrow \pm\infty$. In addition, we will also assume that the rod obeys causality with respect to the time variable. That is, there can be no response in the rod before the heat source is applied. The source term $s(x, t)$ is assumed to be zero for $t < 0$, so that $T(x, t) = 0$ for $t < 0$.

Because Equation 10.276 is a linear equation with homogeneous boundary conditions, it must obey the superposition principle. Thus, if a drive of $s(x, t) = \delta(x - \xi)\delta(t - \tau)$ results in $T(x, t) = g(x|\xi, t|\tau)$, then a drive of

$$s(x, t) = \sum_n \sum_m A_{nm} \delta(x - \xi_n)\delta(t - \tau_m) \tag{10.277}$$

causes a temperature response of

$$T(x, t) = \sum_n \sum_m A_{nm} g(x|\xi_n, t|\tau_m). \tag{10.278}$$

This reasoning can be extended for continuous quantities by converting the summations into integrals. So if the drive is now

$$s(x, t) = \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} d\tau \, s(\xi, \tau)\delta(x - \xi)\delta(t - \tau), \tag{10.279}$$

the response is

$$T(x, t) = \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} d\tau \, s(\xi, \tau) g(x|\xi, t|\tau). \tag{10.280}$$

The Green's function for this problem is the response of the rod to a δ-function impulse at position $x = \xi$ and time $t = \tau$. It obeys the partial differential equation:

$$\frac{\partial g(x|\xi, t|\tau)}{\partial t} - D^2 \frac{\partial^2 g(x|\xi, t|\tau)}{\partial x^2} = \delta(x - \xi)\delta(t - \tau). \tag{10.281}$$

The homogeneous boundary conditions $T(x, t)$ must satisfy are also applied to the Green's function, so $g(x|\xi, t|\tau) \to 0$ as $x \to \pm\infty$. The causality condition requires that $g(x|\xi, t|\tau) = 0$ for $t < \tau$. To solve Equation 10.281, we will take a Fourier transform in space, and a Laplace transform in time. Taking these transforms of the Green's function in two steps gives

$$\underline{G}_F(k|\xi, t|\tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \, g(x|\xi, t|\tau) e^{-ikx} \tag{10.282}$$

$$\underline{G}_{FL}(k|\xi, \underline{s}|\tau) = \int_{\tau}^{\infty} dt \, \underline{G}_F(k|\xi, t|\tau) e^{-\underline{s}t}. \tag{10.283}$$

The Fourier-Laplace transform of Equation 10.281, the original differential equation, gives the algebraic relation

$$\underline{s}\,\underline{G}_{FL}(k|\xi, \underline{s}|\tau) + D^2 k^2 \underline{G}_{FL}(k|\xi, \underline{s}|\tau) = \frac{1}{\sqrt{2\pi}} e^{-ik\xi} e^{-\underline{s}\tau}. \tag{10.284}$$

This allows us to solve for the Fourier-Laplace transform of the Green's function

$$\underline{G}_{FL}(k|\xi, \underline{s}|\tau) = \frac{e^{-ik\xi} \, e^{-\underline{s}\tau}}{\sqrt{2\pi(\underline{s} + D^2 k^2)}}. \tag{10.285}$$

Now all that remains to obtain the Green's function is the double inversion of Equation 10.285. In principle, the inversions could be done in either order. In this case, it is much easier to do the Laplace inversion first.

The Laplace inversion can be evaluated from

$$\underline{G}_F(k|\xi, t|\tau) = \frac{1}{2\pi i} \int_L d\underline{s} \, \frac{e^{-ik\xi} \, e^{-\underline{s}\tau} \, e^{\underline{s}t}}{\sqrt{2\pi(\underline{s} + D^2 k^2)}}. \tag{10.286}$$

There is a single pole at $\underline{s} = -D^2 k^2$. Because the variable k is always on a Fourier contour, it is pure real. The diffusion coefficient D is also pure real. Therefore, this pole is always on the negative, real axis in the complex $\underline{s}$-plane, as shown in Figure 10.29. The Laplace contour is always to the right of this pole. The integral in Equation 10.286 can be performed using closure, closing to the right for $t < \tau$ and to the left for $t > \tau$, with the result

$$\underline{G}_F(k|\xi, t|\tau) = \begin{cases} 0 & t < \tau \\ (1/\sqrt{2\pi})e^{-ik\xi} \, e^{-D^2 k^2(t-\tau)} & t > \tau \end{cases}. \tag{10.287}$$

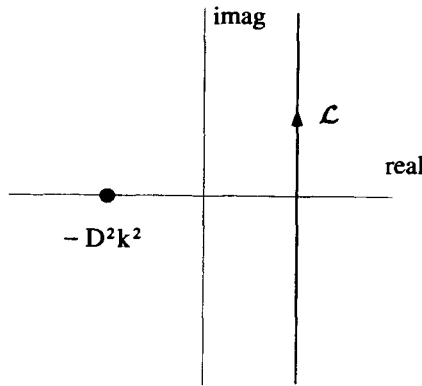Equation 10.287 can be written in more compact form using the Heaviside step function,

Figure 10.29 Laplace Inversion for the Nonhomogeneous Diffusion Equation Green's Funcion

$$H(t - \tau) = \begin{cases} 0 & t < \tau \\ 1 & t > \tau \end{cases}, \qquad (10.288)$$

is

$$\underline{G}_F(k|\xi, t|\tau) = \frac{1}{\sqrt{2\pi}} H(t - \tau) e^{-ik\xi} e^{-D^2 k^2 (t - \tau)}. \qquad (10.289)$$

The space-time Green's function is now obtained by applying a Fourier inversion to Equation 10.289:

$$g(x|\xi, t|\tau) = H(t - \tau) \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \, e^{ik(x - \xi)} e^{-D^2 k^2 (t - \tau)}. \qquad (10.290)$$

One way to solve this integral is by completing the square of the integrand's exponent. Instead, we will use the convolution trick again. Notice the function in Equation 10.289 can be written as the product of two functions:

$$\left[H(t - \tau) \frac{e^{-ik\xi}}{\sqrt{2\pi}} \right] [e^{-D^2 k^2 (t - \tau)}]. \qquad (10.291)$$

Consequently, the Green's function can be written as $1/\sqrt{2\pi}$ times the convolution over x of their inverse transforms:

$$g(x|\xi, t|\tau) = \frac{1}{\sqrt{2\pi}} \left[H(t - \tau) \delta(x - \xi) \otimes \frac{1}{\sqrt{2D^2(t - \tau)}} e^{-x^2/4D^2(t - \tau)} \right]. \qquad (10.292)$$
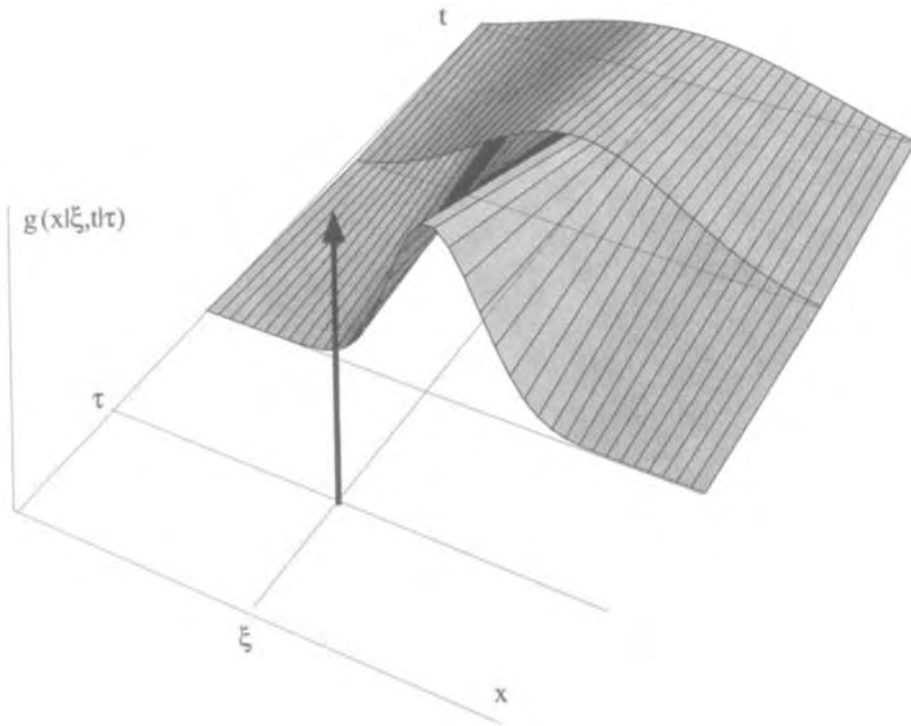
Figure 10.30   Green's Function for the Diffusion Equation with a Distributed Heat Source

Because of the δ-function, this convolution easily evaluates to

$$g(x|\xi, t|\tau) = \frac{H(t - \tau)}{\sqrt{4\pi D^2(t - \tau)}}\, e^{-(x-\xi)^2/4D^2(t-\tau)}. \tag{10.293}$$

A plot of this Green's function is shown in Figure 10.30.

There are some interesting things to notice about this Green's function. First, the $t|\tau$ dependence has the form $t - \tau$ and the $x|\xi$ dependence has the form $x - \xi$. These results make sense both mathematically and physically. Mathematically, this occurs because the problem has constant coefficients and homogeneous boundary conditions at $x = \pm\infty$. Physically, it is clear from symmetry, the problem must be translationally invariant in both time and space. Also notice there is a symmetry with respect to the x and ξ variables, with $g(x|\xi, t|\tau) = g(\xi|x, t|\tau)$, while there is no such symmetry for the t and τ variables, $g(x|\xi, t|\tau) \neq \pm g(x|\xi, \tau|t)$. This symmetry property will be explored in one of the exercises at the end of this chapter.

Example 10.13   As a final example, the Green's function for a driven wave equation will be developed. The one-dimensional driven wave equation for propagation in a

uniform medium can be written as

$$\frac{1}{c_o^2}\frac{\partial^2 u(x,t)}{\partial t^2} - \frac{\partial^2 u(x,t)}{\partial x^2} = s(x,t). \tag{10.294}$$

The constant c_o is the velocity of propagation of the wave. This is a nonhomogeneous equation, with the response $u(x,t)$ generated by the distributed source term $s(x,t)$.

This problem can be treated with a spatial Fourier transform and a temporal Laplace transform, exactly like the diffusion equation of the previous example. There is no problem with this approach. We will demonstrate a different approach, however, because it is one commonly used in physics calculations. To avoid the tedious Laplace transform, we will assume that the time behavior of both the driving term and the response are in the *sinusoidal steady state*, so all the time dependence occurs with a constant amplitude at a fixed frequency ω_o. In other words, we assume that

$$s(x,t) = \text{Real}\left[\underline{S}(x)e^{i\omega_o t}\right], \tag{10.295}$$

and the response has a similar time dependence

$$u(x,t) = \text{Real}\left[\underline{U}(x)e^{i\omega_o t}\right]. \tag{10.296}$$

As you will see, this is very much like taking a Fourier transform in time. Unfortunately, you will also find out it presents a few problems!

In this sinusoidal steady state, the differential equation becomes

$$-\frac{\omega_o^2}{c_o^2}\underline{U}(x) - \frac{d^2\underline{U}(x)}{dx^2} = \underline{S}(x). \tag{10.297}$$

Notice the time dependence completely cancels out, and we have a linear, nonhomogeneous equation for $\underline{U}(x)$. Also, we now are dealing with only a single independent variable, so the partial derivative has been replaced by a regular derivative. The solution for $\underline{U}(x)$ can be formulated in terms of a Green's function,

$$\underline{U}(x) = \int_{-\infty}^{\infty} d\xi\, \underline{S}(\xi)g(x|\xi), \tag{10.298}$$

where $g(x|\xi)$ satisfies the differential equation

$$\left(-\frac{\omega_o^2}{c_o^2} - \frac{d^2}{dx^2}\right)g(x|\xi) = \delta(x-\xi). \tag{10.299}$$

We can take the Fourier transform of both sides of this equation to give

$$\left(-\frac{\omega_o^2}{c_o^2} + k^2\right)\underline{G}(k|\xi) = \frac{1}{\sqrt{2\pi}}e^{-ik\xi}, \tag{10.300}$$
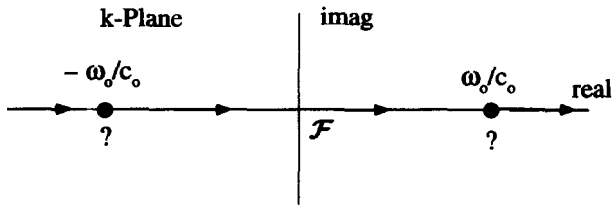
Figure 10.31   Fourier Inversion for the Driven Wave Equation

so that

$$\underline{G}(k|\xi) = \frac{e^{-ik\xi}}{\sqrt{2\pi}(k^2 - \omega_o^2/c_o^2)}. \tag{10.301}$$

The Green's function is then obtained by the Fourier inversion

$$g(x|\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \, \frac{e^{ik(x-\xi)}}{(k + \omega_o/c_o)(k - \omega_o/c_o)}. \tag{10.302}$$

And now you should see the problem. The Fourier contour must remain on the real k-axis, but if it does, it runs directly into the two poles of the integrand! This situation is shown in Figure 10.31. Why did this problem arise, and how do we evaluate the integration near the poles? The answer to the first question is simple. We assumed the response was sinusoidal steady state over time and then carelessly took the Fourier transform over space. This spatial Fourier transform does not exist because the response to a sinusoidal steady-state source located at $x = \xi$ has a finite amplitude for all x from minus to plus infinity. We can't take the Fourier transform of such a function. There are two ways to fix this problem. One approach is to "hand-wave" your way to the correct answer using physical arguments. The other is to modify the problem, so we actually can take the spatial Fourier transform. Both will give the same answer.

The hand-waving approach begins with the observation that what is being described by Equation 10.302 are the waves propagating to the left and to the right from a sinusoidal steady-state source located at $x = \xi$. We will make an attempt to connect the segments of the Fourier contour, shown in Figure 10.31, around the two poles. Then we will check to see if the result makes physical sense, given the physical situation. So let's say the connections are made by small semicircles going above both poles, as shown in Figure 10.32. This contour can be closed in the upper half for $x > \xi$ and in the lower half for $x < \xi$ with the result that

$$g(x|\xi) = \begin{cases} 0 & x > \xi \\ \left[ic_o/(2\omega_o)\right]\left[-e^{i\omega_o(x-\xi)/c_o} + e^{-i\omega_o(x-\xi)/c_o}\right] & x < \xi \end{cases}. \tag{10.303}$$
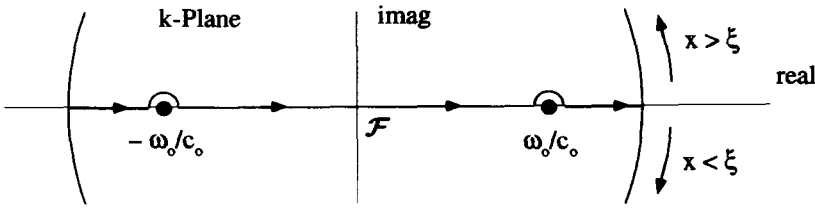
Figure 10.32 First Guess for the Fourier Contour

The first term comes from the pole at $k = \omega_o/c_o$, and the second term comes from the pole at $k = -\omega_o/c_o$.

Remember the Green's function is supposed to be the response to a δ-function source at $x = \xi$. If the sinusoidal steady-state time dependence is put back in the problem, then this solution would say that a source term of

$$s(x, t) = \text{Real}[\delta(x - \xi)e^{i\omega_o t}] \tag{10.304}$$

would give, according to Equation 10.303, a response of

$$u(x, t) = \begin{cases} 0 & x > \xi \\ \text{Real}\left\{\left[i/(2k_o)\right]\left[-e^{i(\omega_o t + k_o(x - \xi))} + e^{i(\omega_o t - k_o(x - \xi))}\right]\right\} & x < \xi \end{cases}, \tag{10.305}$$

where $k_o = \omega_o/c_o$ is a positive constant. There is no response for $x > \xi$, and for $x < \xi$ there are two wave responses, one from the $k = \omega_o/c_o$ pole traveling to the left and one from the $k = -\omega_o/c_o$ pole traveling to the right, as shown in Figure 10.33. The only part of this solution that makes sense is the wave from the $k = \omega_o/c_o$ pole, traveling to the left, away from the source. The other wave in this region, $x < \xi$, is traveling toward the source, which makes no physical sense. It also makes no sense that there is no wave traveling to the right, away from the source in the region $x > \xi$.

It is obvious, however, that all these shortcomings vanish if the "Fourier" contour is modified, as shown in Figure 10.34. Quotation marks are placed around Fourier because, strictly speaking, a Fourier contour exists only on the real axis.
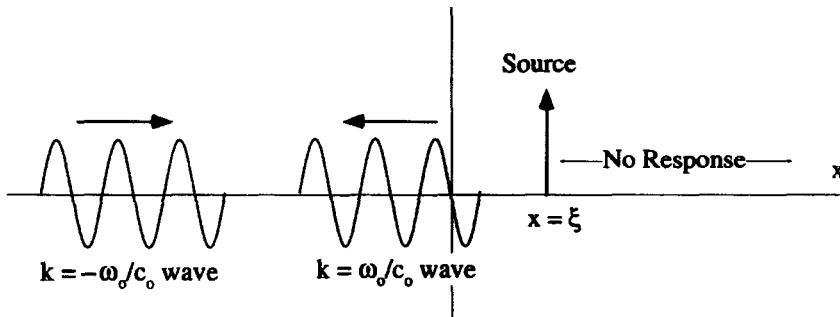

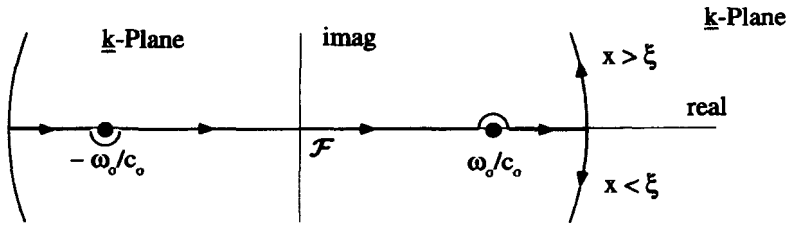**Figure 10.33** Response Using First-Guess Contour

Figure 10.34   Proper Orientation of the "Fourier" Contour

We can reach the same conclusion about how to modify the "Fourier" contour, a little more rigorously, if we introduce a damping term into the original differential equation. If the wave decays as it propagates, we will be able to legitimately take the Fourier transform over x. A simple modification of Equation 10.294 does the trick:

$$\frac{1}{c_o^2}\frac{\partial^2 u(x,t)}{\partial t^2} - \frac{\partial^2 u(x,t)}{\partial x^2} + \alpha\frac{\partial u(x,t)}{\partial t} = s(x,t). \qquad (10.306)$$

Now, if α is positive, the waves are attenuated as they move away from the source. The calculation proceeds as before, except now the damping term modifies the positions of the $\underline{k}$-plane poles of Equation 10.302 and, for small positive α, places them as shown in Figure 10.35. There are no poles along the inversion contour, so everything evaluates easily. Then, finally, we can take the limit as $\alpha \to 0$ to see that the contour of Figure 10.34 is correct. In this limit, the poles in Figure 10.35 become arbitrarily close to the real $\underline{k}$-axis, but the Fourier contour stays below the pole on the right and above the pole on the left. This is exactly the result of the hand-waving discussion that lead to Figure 10.34.

So if we proceed with the inversion of Equation 10.302, using the contour of Figure 10.34 for $\alpha = 0$, we obtain

$$g(x|\xi) = \begin{cases} -\left[i/(2k_o)\right]e^{-ik_o(x-\xi)} & x > \xi \\ -\left[i/(2k_o)\right]e^{+ik_o(x-\xi)} & x < \xi \end{cases}. \qquad (10.307)$$
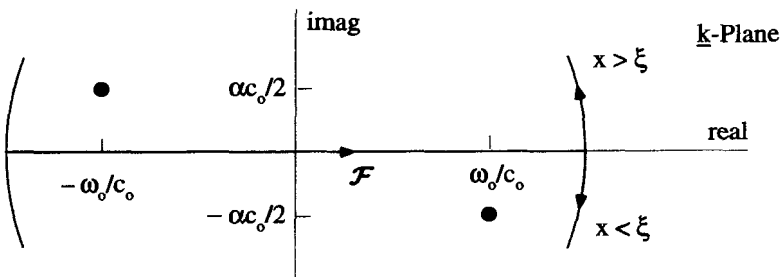

Figure 10.35   Fourier Contour and Complex Plane Picture with Small Wave Damping
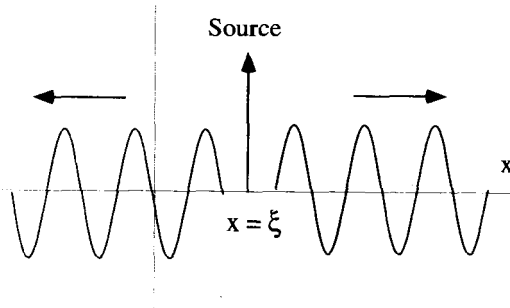
Figure 10.36 Undamped Waves Moving Away from Source

When coupled with the sinusoidal time dependence, this solution behaves as shown in Figure 10.36. This solution describes undamped waves moving in both directions, away from the δ-function source located at $x = \xi$.

For an arbitrary source $\underline{S}(x)$, the response is

$$\underline{U}(x) = \int_{-\infty}^{\infty} d\xi \, \underline{S}(\xi) g(x|\xi) \tag{10.308}$$

$$= \int_{-\infty}^{x} d\xi \, \underline{S}(\xi) \frac{-i}{2k_o} e^{-ik_o(x-\xi)} + \int_{x}^{\infty} d\xi \, \underline{S}(\xi) \frac{-i}{2k_o} e^{ik_o(x-\xi)}.$$

The response to the original problem is obtained by coupling $\underline{U}(x)$ with the sinusoidal steady-state time dependence and taking the real part:

$$u(x, t) = \text{Real}\left[\underline{U}(x) e^{i\omega_o t}\right]. \tag{10.309}$$

The Green's function solution for this problem, Equation 10.308, is easier to understand if it is assumed that $\underline{S}(x)$ is pure real. Then

$$s(x, t) = \text{Real}\left[S(x) e^{i\omega_o t}\right]$$

$$= S(x) \cos(\omega_o t), \tag{10.310}$$

and the response becomes

$$u(x, t) = \int_{-\infty}^{x} d\xi \, S(\xi) \frac{1}{2k_o} \sin\left[\omega_o t - k_o(x - \xi)\right]$$

$$+ \int_{x}^{\infty} d\xi \, S(\xi) \frac{1}{2k_o} \sin\left[\omega_o t + k_o(x - \xi)\right]. \tag{10.311}$$

From this expression, it can be seen that the response is a sum of waves propagating away from the distributed source, toward the observation point x.

EXERCISES FOR CHAPTER 10

1. Determine whether the following differential equations are homogeneous or nonhomogeneous, and linear or nonlinear:

   i. $\dfrac{d}{dx}\left[\dfrac{dy(x)}{dx}\right]^2 - 2y(x)\dfrac{dy(x)}{dx} = 0$

   ii. $\dfrac{d}{dx}\left[\dfrac{dy(x)}{dx}\right]^2 - 2y(x)\dfrac{dy(x)}{dx} = x$

   iii. $x^2\dfrac{d^2y(x)}{dx^2} + x\dfrac{dy(x)}{dx} + x = 0$

   iv. $x^2\dfrac{d^2y(x)}{dx^2} + x\dfrac{dy(x)}{dx} + \dfrac{1}{y(x)} = 0$

2. The rate of evaporation for a constant-density, spherical drop is proportional to its surface area. Assume this evaporation is the only way the drop can lose mass and determine a differential equation that describes the time dependence of the radius of the drop. Solve this equation for the radius as a function of time, assuming that at $t = 0$ the radius is r_o.

3. Using separation of variables, solve the differential equation

   $$\frac{dy(x)}{dx} = x^2 y^2,$$

   with the boundary condition $y(0) = 5$. What happens when you try to impose the boundary condition that $y(0) = 0$?

4. Consider the differential equation

   $$\frac{dy(x)}{dx} = -\frac{y}{x},$$

   with the boundary condition $y(1) = 1$.
   (a) Solve this equation using separation of variables.
   (b) This equation can also be solved by converting it into an exact differential. Using the notation of Equations 10.21–10.24, identify the functions $R(x, y)$, $S(x, y)$, and $\phi(x, y)$. Using the boundary condition, find a solution $\phi(x, y) = C_o$ that is equivalent to your answer in part (a).

5. Consider the differential equation

   $$\frac{dy(x)}{dx} + \frac{y(x)}{x} = \cos(x^2).$$

   (a) Determine an integrating factor for this equation.
   (b) Solve for $y(x)$.

6. A resistor R and capacitor C are driven by a voltage source $v(t)$, as shown in the circuit below:
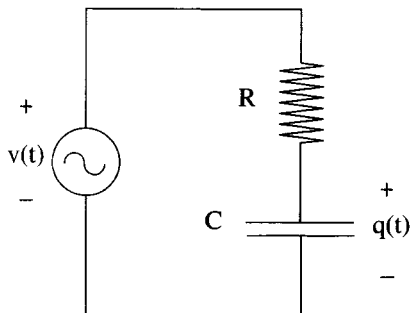
The charge on the capacitor is given by the function $q(t)$. For $t < 0$, the voltage source is zero and the charge on the capacitor is zero. The differential equation that governs the charge is

$$R\frac{dq(t)}{dt} + \frac{1}{C}q(t) = v(t).$$

(a) Find the integrating factor for this differential equation.

(b) Using the integrating factor, express $q(t)$ in terms of $v(t)$.

(c) Solve the equation of part (b) above if $v(t) = \delta(t)$.

(d) Solve the equation of part (b) above if $v(t)$ is a unit step applied at $t = 0$.

7. Find the integrating factor and the general solution, with no specific boundary condition, for the differential equation

$$\frac{1}{y(x)}\frac{dy(x)}{dx} + \cot x = \frac{\cos^2 x}{y(x)}.$$

8. Consider the first-order differential equation

$$\frac{dy(x)}{dx} - x\,y(x) = x.$$

(a) What is the integrating factor for this differential equation?

(b) Find $y(x)$ subject to the boundary condition $y(0) = 0$.

9. Consider Bessel's equation of order zero:

$$x\frac{d^2 y(x)}{dx^2} + \frac{dy(x)}{dx} + x\,y(x) = 0.$$

Show that the Wronskian for this homogeneous differential equation has the form

$$W(x) = \frac{C_o}{x},$$

where C_o is a constant.

10. What is the Wronskian associated with the differential equation

$$\frac{d^2 y(x)}{dx^2} + x \frac{dy(x)}{dx} + \frac{1}{x^2} y(x) = 0?$$

11. What is the Wronskian associated with the differential equation

$$\frac{d^2 y(x)}{dx^2} + y(x) \sin(\omega_o x) = 0?$$

12. Given that one solution of

$$\frac{d^2 y(x)}{dx^2} + \frac{1}{x} \frac{dy(x)}{dx} - \frac{m^2}{x^2} y(x) = 0$$

is $y(x) = x^m$, find the second solution. Using your result above, find a particular solution to the nonhomogeneous equation

$$x^2 \frac{d^2 y(x)}{dx^2} + x \frac{dy(x)}{dx} - y(x) = \frac{1}{1 - x}.$$

13. Use the method of Frobenius to find a series solution to

$$x^2 \frac{d^2 y(x)}{dx^2} + 4x \frac{dy(x)}{dx} + x y(x) = 0.$$

14. The Legendre equation has the form

$$(1 - x^2) \frac{d^2 y(x)}{dx^2} - 2x \frac{dy(x)}{dx} + n(n + 1)y(x) = 0.$$

Obtain the indicial equation used in the method of Frobenius to find a series representation for $y(x)$.

15. Show that the differential equation

$$x \frac{d^2 y(x)}{dx^2} + (1 - x) \frac{dy(x)}{dx} + \lambda y(x) = 0$$

has Frobenius polynomial solutions, as long as λ is a nonnegative integer. Determine the solution for $\lambda = 4$ and with the boundary condition $y(0) = 1$.

16. The differential equation

$$\frac{d^2y(x)}{dx^2} - x\,y(x) = 0,$$

possesses two independent solutions that are called the Airy functions. Use the method of Frobenius to evaluate these functions. What is the Wronskian of this differential equation?

17. Using the method of Frobenius solve the differential equation

$$x^2\frac{d^2y(x)}{dx^2} + (x^2 + x)\frac{dy(x)}{dx} + y(x) = 0.$$

18. Use the method of Frobenius to obtain two independent solutions to the equation

$$2x^2\frac{d^2y(x)}{dx^2} + 2x^2\frac{dy(x)}{dx} + y(x) = 0.$$

Show that your solutions converge and prove they are independent.

19. Use the method of Frobenius to generate a series solution for the differential equation

$$\frac{d^3y(x)}{dx^3} - x\frac{dy(x)}{dx} = 0.$$

Determine the generating function for the coefficients of the series and write out the first few nonzero terms.

20. A tank car with a hole in the bottom is dumping its load so its total mass as a function of time is given by $m(t) = m_o(1 - \alpha_o t)$, where α_o and m_o are constants. The car sits on a frictionless track and is attached to a spring with a spring constant K_o.
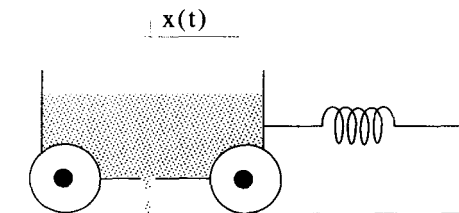
The car is released at $t = 0$ from $x = 0$ with an initial velocity v_o. Assuming the spring is at equilibrium when the car is at $x = 0$,

(a) what is the second-order, homogeneous differential equation that describes the motion of the car?

(b) is this a linear or nonlinear differential equation?

(c) use the method of Frobenius to solve for $x(t)$ over the interval $0 < t < 1/\alpha_o$.

(d) check the convergence of your solution.

(e) what happens at $t = 1/\alpha_o$?

21. The quantum mechanical interaction between two nuclear particles can be described by the one-dimensional interaction potential

$$V(x) = -\frac{e^{-\alpha x}}{x}.$$

Develop a series solution for the wave function $\Psi(x)$ that describes this system. $\Psi(x)$ satisfies the Schrödinger equation,

$$\frac{\hbar^2}{2m} \frac{d^2\Psi(x)}{dx^2} + [E - V(x)]\,\Psi(x) = 0,$$

where E is a constant representing the energy of the particle and m is its mass. Explicitly write out the first three terms of the series you obtain.

22. Consider an undamped, undriven, frictionless mass spring system described by the homogeneous differential equation

$$m\frac{d^2y(t)}{dt^2} + K\,y(t) = 0,$$

where m is the mass and K is the spring constant. For this problem we will allow the spring to get stiffer as it is compressed. To accomplish this, let the spring constant K become dependent on the displacement such that

$$K = K_o y^2,$$

where K_o is a positive constant.

(a) At $t = 0$, the mass is released from $x = x_o$ with no initial velocity. Using the differential equation describing the motion, obtain an expression relating the velocity of the mass to its position. What is the acceleration of the mass when $x = x_o$?

(b) Make a graph of the velocity vs. position to show how the mass oscillates.

23. A mass spring oscillator is set up with a magic spring, with a negative spring constant. It is frictionless and driven by an external force $f(t)$ that is a function of time. The force from the spring on the mass is given by $K_o x(t)$ rather than $-K_o x(t)$. The position of the mass is $x(t)$ and the initial conditions are that $x = 0$ and $dx/dt = 0$ for all $t < 0$.

(a) Write down the differential equation of motion for $x(t)$, the position of the mass.

(b) Let the external force $f(t)$ be given by

$$f(t) = \begin{cases} f_o \sin \omega_o t & t > 0 \\ 0 & t < 0 \end{cases}.$$

What is the Laplace transform of $x(t)$?

(c) What is $x(t)$ for large t?

24. A driven, linear mass spring oscillator is described by the differential equation

$$m \frac{d^2 x(t)}{dt^2} + Kx(t) = f(t),$$

where $x(t)$ is the position of the mass as a function of time, and $f(t)$ is the driving force

$$f(t) = \begin{cases} t e^{-t} & t > 0 \\ 0 & t < 0 \end{cases}.$$

For $t < 0$ the mass is at rest.

(a) If $x(t)$ has a Laplace transform $X(\underline{s})$, what is the Laplace transform of $d^2 x(t)/dt^2$?

(b) What is $x(t)$ for all $t > 0$?

25. A linear, frictionless mass-spring oscillator with mass m and spring constant K is at rest for $t < 0$ and is driven by an external force $f(t)$.

(a) If $f(t) = \delta(t)$, what initial conditions are established just after this δ-function force is applied?

(b) If $f(t) = d\delta(t)/dt$, what are these initial conditions?

(c) Finally, find the position and velocity of the mass for all $t \geq 0$ if

$$f(t) = \begin{cases} \delta(t) + t e^{-t} & t \geq 0 \\ 0 & t < 0 \end{cases}.$$

26. Solve the nonlinear differential equation

$$[y(x)]^{1/2} \frac{d^2 y(x)}{dx^2} = 1/3,$$

given the boundary conditions $y = 0$ and $dy/dt = 0$ at $x = x_o$.

27. Find the unique solution for $y(x)$ to the nonlinear equation

$$\frac{d^2 y(x)}{dx^2} = \frac{3}{4} [y(x)]^{-1/3},$$

with the boundary conditions $y(0) = 0$ and $dy/dx|_0 = 0$.

28. Use the method of quadrature to solve the equation

$$\frac{d^2 y(x)}{dx^2} - [y(x)]^3 = 0,$$

if $y(0) = 1$ and $dy/dx|_0 = 1/\sqrt{2}$.

29. Use the method of quadrature to solve the nonlinear equation

$$\frac{d^2 y(x)}{dx^2} = -\cos y \sin y,$$

given the boundary conditions $y(0) = 0$ and $dy/dx|_0 = 1$.

30. Use the method of quadrature to solve the differential equation

$$\frac{d^2 y(x)}{dx^2} = \frac{\sin y}{\cos^3 y}$$

for $y(x)$. Show, by substituting it back into the equation, that your solution works. Choose the constants of integration to give the simplest possible answer.

31. This problem involves a Child–Langmuir solution for a two-component, anode-cathode gap. The geometry is described in the figure below. A large source of electrons exists at the cathode, and we assume the cathode temperature is sufficient for the electric field to be zero at $t = 0$. The same is true at the anode, except the particles there are positively charged ions. Let the electron have a mass of m_- and a charge $-q_o$, while the ion has a mass m_+ and charge $+q_o$.
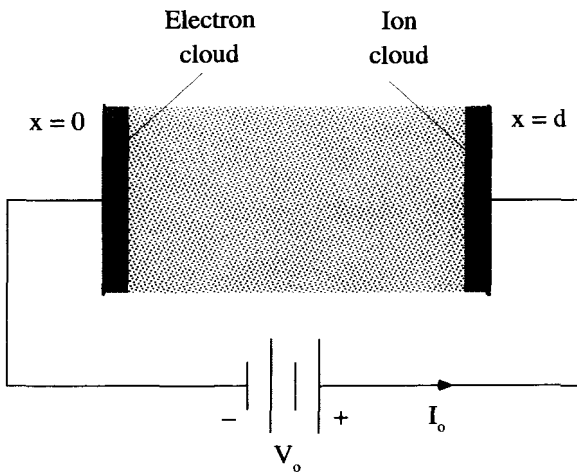
Look for a steady-state solution such that $\partial/\partial t$ of all quantities are zero. Both electrons of charge density $\rho_-(x)$ and ions of charge density $\rho_+(x)$ will be present

in the space between the electrodes, so the relevant Maxwell equation becomes

$$\overline{\nabla} \cdot \epsilon_o \overline{E}(x) = \rho_+(x) + \rho_-(x).$$

Use the method of quadrature to obtain a relationship between the current I_o and the voltage V_o.

This problem is quite involved. Here are two hints to get you started:

(1) The example worked out in this chapter arrived at the equation:

$$\frac{1}{2}\frac{d}{dx}\left[\frac{d\phi(x)}{dx}\right]^2 = \frac{J_o}{\epsilon_o}\sqrt{\frac{m}{2q_o}}\,\phi^{-1/2}(x)\,\frac{d\phi(x)}{dx}.$$

The solution to this exercise works out best if an indefinite integral, with its constants of integration, is used to impose the boundary conditions, instead of the definite integral used in the example worked out in this chapter. Proceed as follows:

$$\int d\left[\frac{1}{2}\left(\frac{d\phi}{dx}\right)^2\right] = \frac{J_o}{\epsilon_o}\sqrt{\frac{m}{2q_o}}\int d\phi\,\phi^{-1/2}$$

$$\frac{1}{2}\left(\frac{d\phi}{dx}\right)^2 + C_1 = \frac{J_o}{\epsilon_o}\sqrt{\frac{m}{2q_o}}\,2\,\phi^{1/2} + C_2.$$

The two constants can be combined

$$\frac{1}{2}\left(\frac{d\phi}{dx}\right)^2 = \frac{J_o}{\epsilon_o}\sqrt{\frac{m}{2q_o}}\,2\,\phi^{1/2} + C_o,$$

and the single constant C_o evaluated by the boundary conditions,

$$\phi(0) = 0, \qquad \left.\frac{d\phi}{dx}\right|_{x=0} = 0.$$

For the example worked out in this chapter, $C_o = 0$. For this exercise, this constant will not be zero.

(2) In the chapter, we ended up with an integral in the form

$$\int_0^{V_o} d\phi\,\frac{1}{\phi^{1/4}}.$$

This integral was easy to perform. This problem does not work out as nicely, and the integral must be done numerically. For numerical integrations, variables are usually normalized so the results will be more general and can be easily scaled. For example, to normalize the above integral, we let

$$\phi(x) = V_o\eta(x),$$

so that the normalized form becomes

$$V_o^{3/4} \left[\int_0^1 d\eta \, \frac{1}{\eta^{1/4}} \right].$$

The quantity in the brackets is referred to as a normalized integral because the variable η is dimensionless and the limits are from zero to one. The normalized integral for this exercise should come out to be

$$\int_0^1 d\eta \, \frac{1}{[\eta^{1/2} + (1 - \eta)^{1/2} - 1]^{1/2}}.$$

Done numerically this integral has a value of about $(4/3)\sqrt{1.86}$.

32. For the string problem worked out in this chapter, we found that the Green's function solution took the form

$$y(x) = \int_a^x d\xi \, f(\xi) g_2(x|\xi) + \int_x^b d\xi \, f(\xi) g_1(x|\xi).$$

Show that the derivative of $y(x)$ can be expressed as

$$\frac{dy(x)}{dx} = \int_a^x d\xi \, f(\xi) \frac{dg_2(x|\xi)}{dx} + \int_x^b d\xi \, f(\xi) \frac{dg_1(x|\xi)}{dx}.$$

33. For the string problem worked out in this chapter, the displacement of the string was governed by the differential equation

$$T \frac{d^2y(x)}{dx^2} = f(x).$$

Assume the homogeneous boundary conditions $y(0) = y(L) = 0$, and use the Green's function to find the displacement of the string if

$$f(x) = F_o \left[(x - L/2)^2 - L^2/4 \right] \qquad 0 < x < L.$$

34. Derive the Green's function for the string problem again, using the boundary conditions
(a) $y(0) = y_1$; $y(L) = y_2$.
(b) $y(0) = 0$; $dy/dx|_L = 0$.
(c) $y(0) = 0$; $dy/dx|_L = A_o$.
Show, with drawings, how the string can be fastened at its end points to impose the three different boundary conditions described above. Why do the boundary conditions

$$dy/dx|_0 = 0; \quad dy/dx|_L = 0$$

cause problems with the Green's function derivation?

35. Poisson's equation can be viewed as a multidimensional form of the string problem. Two spatial derivatives of the dependent variable generate a function proportional to the drive

$$\nabla^2 \Phi(\bar{r}) = -\frac{\rho_c(\bar{r})}{\epsilon_o}.$$

In this expression, $\Phi(\bar{r})$ is the electrostatic potential generated by a charge distribution $\rho(\bar{r})$.

(a) Express a point charge in two dimensions using polar coordinates ρ and ϕ.

(b) Find the Green's function for the two-dimensional Poisson's equation using the polar coordinates. What boundary condition does this Green's function satisfy?

(c) Using your Green's function, find a solution for $\Phi(\bar{r})$ for an arbitrary $\rho_c(\bar{r})$.

(d) Repeat the above steps for a three-dimensional solution using spherical coordinates.

(e) Using your answer to part (d) above, find the solution for the potential of the dipole charge distribution shown below, where $\pm q$ are point charges separated by a distance d.
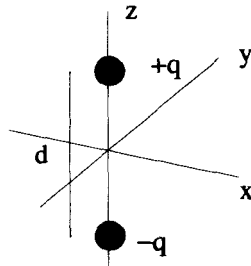
36. If time dependence is included in the description of an infinitely long, stretched string, the differential equation for its displacement becomes a wave equation

$$\frac{\partial^2 y(x,t)}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 y(x,t)}{\partial t^2} = 0.$$

The string is stretched in the x-direction, and this equation describes perpendicular displacements in the y-direction. If there are no driving forces along the string, the differential equation is homogeneous, and excitations are established by the initial conditions at $t = 0$, which then propagate for $t > 0$.

(a) Assume that the initial conditions are $y(x, t = 0) = y_o(x)$ and $\partial y/\partial t|_{t=0} = 0$. Find the Green's function that allows the displacement of the string to be written as

$$y(x,t) = \int_{-\infty}^{\infty} d\xi \, y_o(\xi) g(x|\xi, t).$$

(b) Plot your solution for the Green's function vs. x for three values of t.

(c) Find $y(x, t)$ if

$$y_o(x) = \begin{cases} 1 & -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}.$$

(d) Now find the Green's function for the initial conditions $y(x, t = 0) = 0$ and $\partial y/\partial t|_{t=0} = v_o(x)$. Plot this Green's function vs. x for three different values of t.

37. A linear, driven, mass-spring oscillator is acted upon by a standard $-Kx$ spring force, an external force $f(t)$, and a magic antifriction force $\alpha\, dx/dt$ where α is small. Assume the mass is at rest for $t < 0$ and its position is given by $x(t)$.

(a) Write down the differential equation of motion for this system.

(b) Take $f(t)$ to be zero for $t < 0$ and equal to $F_o \sin(\omega_o t)$ for $t > 0$. Use Laplace transform techniques to obtain a solution for $x(t)$ as $t \to \infty$.

(c) Sketch the Green's function $g(t - \tau)$ vs. τ.

(d) If $f(t) = t$ for $t > 0$, what is $x(t)$?

38. Consider the differential equation

$$\frac{d^3 r(t)}{dt^3} + r(t) = d(t),$$

where $r(t)$ is the response to a drive $d(t)$. The drive and response are assumed to be zero for $t < 0$.

(a) What is the Green's function for this differential equation? Can it be obtained with Fourier transform techniques? Can it be obtained with Laplace transform techniques? Discuss the initial conditions on $r(t)$ and its derivatives.

(b) Sketch the Green's function $g(t|\tau)$ vs. t.

(c) If $d(t) = t$ for $t > 0$, what is the response $r(t)$?

39. Consider the differential equation

$$\frac{d^4 r(t)}{dt^4} - r(t) = s(t),$$

where $r(t)$ is the response to a source $s(t)$, and the independent variable t represents time. Since this is a fourth-order differential equation, four boundary conditions are necessary to obtain a unique solution. For all parts of this problem assume that $r(t)$ and $s(t)$ are zero for $t < 0$.

(a) What initial conditions are established if $s(t) = \delta(t)$? Assume these conditions are imposed just after the δ-function occurs.

(b) Obtain an expression for the Green's function $g(t|\tau)$ for this problem.

(c) Make a plot of $g(t|\tau = 0)$ vs t.

40. Find the Green's function for the differential equation

$$\frac{d^2y(x)}{dx^2} + y(x) = f(x)$$

over the interval $0 < x < 1$, with the boundary conditions $y(0) = y(1) = 0$. Solve for $y(x)$ if $f(x) = \sin(\pi x)$.

41. Consider the linear operator

$$\mathcal{L}_{op} = \frac{d^2}{dx^2} + 1$$

and the differential equation

$$\mathcal{L}_{op}\, y(x) = f(x).$$

For the initial conditions set $y(0) = 0$ and $dy/dx|_{x=\pi} = 0$.

(a) What is the Green's function $g(x|\xi)$ for this operator and what are its boundary conditions?

(b) Obtain the Green's function solution for $y(x)$ as specifically as you can, while still leaving $f(x)$ arbitrary.

(c) Now set $f(x) = \delta(x - x_o)$ and, using your answer to part (b), determine $y(x)$.

42. The equation of motion for a damped harmonic oscillator, driven by a force $f(t)$, takes the form

$$m\frac{d^2x(t)}{dt^2} + \alpha\frac{dx(t)}{dt} + K x(t) = f(t)$$

where the damping constant α is greater than zero, and K is the positive spring constant. Assume the drive and response are zero for $t < 0$.

(a) Use Laplace transform techniques to obtain the Green's function for this problem.

(b) What is $x(t)$ if

$$f(t) = \begin{cases} F_o & 0 < t < T_o \\ 0 & \text{otherwise} \end{cases} ?$$

43. Consider the following circuit, where $i_s(t)$ is a current source, which is zero for $t < 0$. Assume that the response $v_r(t)$, which is the voltage across the capacitor, is also zero for $t < 0$.
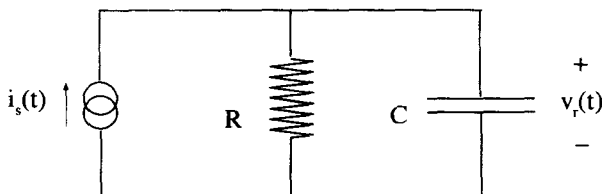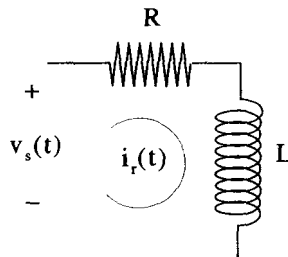
(a) Obtain a nonhomogeneous differential equation which relates the response $v_r(t)$ to the drive $i_s(t)$.

(b) Working in the time domain, determine the Green's function for this problem.

(c) Using your Green's function form part (b), determine $v_r(t)$ if

$$i_s(t) = \begin{cases} I_0 t^2 & t > 0 \\ 0 & t < 0 \end{cases}.$$

(d) Now find the solution working in the frequency domain. Obtain the Laplace transform of the response $\underline{V}_r(s)$ in terms of the Laplace transform of the drive $\underline{I}_s(s)$. Use this result to obtain $v_r(t)$ for the drive given in part (c), above. Show that you obtain the same result as given by the Green's function solution.

(e) Show that the relationship between $\underline{V}_r(s)$ and $\underline{I}_s(s)$ can be obtained by treating the complex impedance of the capacitor as $1/(C\underline{s})$ and forming the parallel combination with the impedance of the resistor R.

(f) Using the complex impedance idea of part (e), what is the complex impedance of an inductor, L? What is the ratio $\underline{V}(s)/\underline{I}(s)$ for the following circuit?
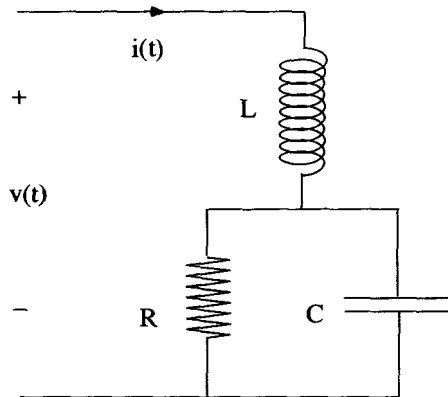
44. Consider the circuit, where $v_s(t)$ is a source voltage, and the current $i_r(t)$ is looked at as the response.

(a) Obtain the nonhomogeneous differential equation that describes the response $i_r(t)$ in terms of the source $v_s(t)$.

(b) Working in the time domain, determine the Green's function for this differential equation. Assume that the source and response are zero for $t < 0$.

(c) Using your Green's function from part (b), determine $i_r(t)$ if

$$v_s(t) = \begin{cases} V_o t & t > 0 \\ 0 & t < 0 \end{cases}.$$

(d) Now find the solution working in the frequency domain. Obtain $\underline{I}_r(\underline{s})$, the Laplace transform of $i_r(t)$, in terms of $\underline{V}_s(\underline{s})$, the Laplace transform of $v_s(t)$. Use the same function for $v_s(t)$ as was used in part (c). Invert the Laplace transform for $i_r(t)$ to obtain the same answer you got using the Green's function approach of part (c).

(e) Show that by considering the impedance of the inductor to be $L\underline{s}$, the relationship between $\underline{I}_r(\underline{s})$ and $\underline{V}_s(\underline{s})$ can be obtained by inspection. Notice that this is just the Laplace transform of the impulse response.

45. Using the Green's function technique, solve the differential equation

$$\frac{d^2 y(x)}{dx^2} - k_o^2\, y(x) = f(x)$$

over the interval $0 < x < L_o$, subject to the boundary conditions $y(0) = y(L_o) = 0$, using the following steps:

(a) Show that the Green's function for $x < \xi$ is given by

$$g(x|\xi) = -\frac{\sinh(k_o x)\sinh(k_o L_o - k_o \xi)}{k_o \sinh(k_o L_o)}.$$

(b) Find the Green's function for $x > \xi$.

(c) Find $y(x)$ in terms of $f(x)$ using the Green's function integral.

(d) Find $y(x)$ under the above conditions if

$$f(x) = 5\delta(x - 5) + 10\delta(x - 10).$$

(e) How does your answer to part (c) change if $y(0) = 0$ and $y(L_o) = 5$?

46. Consider the differential equation

$$\frac{d}{dx}\left[x\,\frac{dy(x)}{dx}\right] = s(x)$$

over the interval $0 < x < 1$, with $s(x)$ a known function of x. The solution $y(x)$ is to satisfy the following boundary conditions

$$y(0) \to \text{finite} \quad ; \quad y(1) = 0.$$

(a) What differential equation must be satisfied by $g(x|\xi)$, the Green's function for this problem?

(b) What is the general form of $g(x|\xi)$ for $x < \xi$ and $x > \xi$?

(c) What conditions must $g(x|\xi)$ and its derivative with respect to x satisfy at $x = \xi$?

(d) Using your results from part (c) and the boundary conditions, determine $g(x|\xi)$ in the range $0 < x < 1$.

(e) Sketch $g(x|\xi)$ vs. x for a fixed ξ, over the range $0 < x < 1$.

47. Find the Green's function for the differential equation

$$x^2 \frac{d^2 y(x)}{dx^2} + x \frac{dy(x)}{dx} - y(x) = d(x)$$

subject to the boundary conditions $y(0) = y(1) = 0$. The Green's function must satisfy

$$\left[x^2 \frac{d^2}{dx^2} + x \frac{d}{dx} - 1 \right] g(x|\xi) = \delta(x - \xi).$$

It may be easier, however, to solve for the Green's function from

$$\left[x \frac{d^2}{dx^2} + \frac{d}{dx} - \frac{1}{x} \right] g(x|\xi) = \frac{1}{x} \delta(x - \xi).$$

48. This problem involves the diffusion of magnetic field into an electrically conducting material. The material is characterized by a constant electrical conductivity σ_o and magnetic susceptibility μ_o and exists in the half-plane $x > 0$. Free space fills the region $x < 0$.
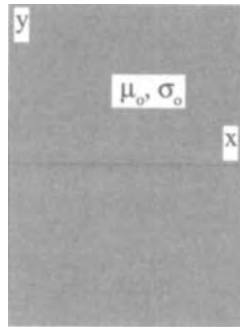
The electric $\overline{\mathbf{E}}$ and the magnetic $\overline{\mathbf{B}}$ fields in the conducting material are governed by Maxwell's equations in the diffusion limit

$$\overline{\nabla} \times \overline{\mathbf{B}} = \sigma_o \mu_o \overline{\mathbf{E}}$$

$$\overline{\nabla} \times \overline{\mathbf{E}} = -\frac{\partial \overline{\mathbf{B}}}{\partial t}$$

$$\overline{\nabla} \cdot \overline{\mathbf{B}} = 0.$$

Treat the material as if it were one-dimensional (i.e., all quantities have $\partial/\partial z = 0$ and $\partial/\partial y = 0$) and use these equations to obtain the diffusion equation

$$\frac{\partial^2 \overline{\mathbf{B}}(x,t)}{\partial x^2} = \sigma_o \mu_o \frac{\partial \overline{\mathbf{B}}(x,t)}{\partial t}.$$

Now consider the situation where the magnetic field is controlled in the free space region so that $\overline{\mathbf{B}}(x,t) = B_o(t)\hat{\mathbf{e}}_y$ for $x < 0$. A Green's function solution for $\overline{\mathbf{B}}(x,t)$ in the region $x \geq 0$ is sought, so that the solution for the magnetic field can be placed in the following form:

$$\overline{\mathbf{B}}(x,t) = \int_{-\infty}^{\infty} d\tau \, B_o(\tau) g(x,t|\tau)\hat{\mathbf{e}}_y.$$

(a) What differential equation and what boundary conditions at $x = 0$ must $g(x,t|\tau)$ satisfy?

(b) Determine $\underline{G}(x, \omega|\tau)$, the Fourier transform of $g(x,t|\tau)$.

(c) Describe the restrictions that must be placed on the function $B_o(t)$ in order for the above solution to be valid.

49. Consider the one-dimensional wave equation

$$\left[\frac{\partial^2}{\partial x^2} - \frac{1}{c_o^2}\frac{\partial^2}{\partial t^2}\right] r(x,t) = s(x,t),$$

where the source term $s(x,t)$ and the response $r(x,t)$ are both zero for $t < 0$.

(a) Obtain the Green's function $g(x|\xi, t|\tau)$ for this problem that allows the response to be written as

$$r(x,t) = \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} d\tau \, s(\xi, \tau) g(x|\xi, t|\tau).$$

(b) Plot $g(x|\xi, t|\tau)$ vs. x for three different values of t, one value less than τ and two values greater than τ.

Note that if $f(x)$ is a unit step from $-a < x < a$, the Fourier transform of $f(x)$ is

$$f(x) \leftrightarrow \sqrt{\frac{2}{\pi}} \frac{\sin(ka)}{k}.$$

50. In the chapter, we derived the Green's function for the wave equation in Equation 10.307.

   (a) How is $g(x|\xi)$ modified if we have the additional boundary condition $g(x = 0|\xi) = 0$?

   (b) Make a plot of this Real$[g(x|\xi)e^{i\omega_o t}]$ for several different times—as many as are needed to show its character.

   (c) What physical situation does this new boundary condition represent?

51. In free space, electromagnetic waves propagate according to Maxwell's equations

$$\overline{\nabla} \times \overline{\mathbf{E}}(\overline{\mathbf{r}}, t) = -\frac{\partial \mu_o \overline{\mathbf{H}}(\overline{\mathbf{r}}, t)}{\partial t}$$

$$\overline{\nabla} \times \overline{\mathbf{H}}(\overline{\mathbf{r}}, t) = +\frac{\partial \epsilon_o \overline{\mathbf{E}}(\overline{\mathbf{r}}, t)}{\partial t},$$

which can be combined to give the wave equation

$$-\nabla^2 \overline{\mathbf{E}}(\overline{\mathbf{r}}, t) + \frac{\partial^2 \mu_o \epsilon_o \overline{\mathbf{E}}(\overline{\mathbf{r}}, t)}{\partial t^2} = 0.$$

If, instead of free space, the wave is propagating in a dielectric medium that is spatially dependent, i.e.,

$$\epsilon_o \rightarrow \epsilon_o + \epsilon(\overline{\mathbf{r}}),$$

   (a) Show that the wave equation becomes

$$-\nabla^2 \overline{\mathbf{E}}(\overline{\mathbf{r}}, t) + \frac{\partial^2 \mu_o \epsilon_o \overline{\mathbf{E}}(\overline{\mathbf{r}}, t)}{\partial t^2} + \frac{\partial^2 \mu_o \epsilon(\overline{\mathbf{r}}) \overline{\mathbf{E}}(\overline{\mathbf{r}}, t)}{\partial t^2} = 0.$$

   (b) Consider a wave propagating at a pure real frequency ω_o, so that $\overline{\mathbf{E}}(\overline{\mathbf{r}}, t) \rightarrow \underline{\overline{\mathbf{E}}}(\overline{\mathbf{r}}) e^{i\omega_o t}$. This is like taking a Fourier transform in time. Show that the wave equation becomes

$$-\nabla^2 \underline{\overline{\mathbf{E}}}(\overline{\mathbf{r}}) - \omega_o^2 \mu_o \epsilon_o \underline{\overline{\mathbf{E}}}(\overline{\mathbf{r}}) = +\omega_o^2 \mu_o \epsilon(\overline{\mathbf{r}}) \underline{\overline{\mathbf{E}}}(\overline{\mathbf{r}}).$$

Now let $\epsilon(\overline{\mathbf{r}}) = \delta(x)$ to model a sheet of infinite dielectric in the $x = 0$ plane. Assume there is no variation of the electric field in the y-direction, and that the spatial dependence of the field can be expressed as

$$\underline{\overline{\mathbf{E}}}(\overline{\mathbf{r}}) = \underline{E}_o \, e^{ik_x x} \, e^{ik_z z} \, \hat{\mathbf{e}}_y.$$

This is like taking Fourier transforms in the two spatial directions.

   (c) Find k_x and k_z for $x > 0$ and $x < 0$.

   (d) Can this solution for the electric field be used as a Green's function for finding the solution for an arbitrary $\epsilon(x)$?

52. This problem looks at the diffusion of heat along an infinite rod with the temperature at $x = 0$ controlled for all time,

$$T(x = 0, t) = \begin{cases} f(t) & t > 0 \\ 0 & t < 0 \end{cases}.$$

The standard diffusion equation applies

$$D^2 \frac{\partial^2 T(x,t)}{\partial x^2} - \frac{\partial T(x,t)}{\partial t} = 0,$$

which, with the above boundary condition and the assumption that $T(x,t) \to 0$ as $|x| \to \infty$, uniquely determines $T(x,t)$ for all x and $t > 0$. The solution for $T(x,t)$ can be obtained from a Green's function integral

$$T(x,t) = \int_0^\infty d\tau \, f(\tau) \, g(x,t|\tau).$$

(a) Identify the differential equation and boundary conditions that $g(x,t|\tau)$ must satisfy.

(b) Determine $g(x,t|\tau)$. The following Laplace transform will probably be of help:

$$\frac{k}{2\sqrt{\pi t^3}} e^{-k^2/(4t)} \leftrightarrow e^{-k\sqrt{s}} \qquad (k > 0).$$

53. A differential equation is said to be in Sturm-Liouville form if it can be written as

$$\frac{d}{dx}\left[p(x) \frac{dy(x)}{dx} \right] + q(x) \, y(x) = s(x),$$

where $a(x)$, $b(x)$, and $s(x)$ are known functions of the independent variable. All linear, second-order differential equations can be placed in this form. This particular form has many advantages, as will be shown in the next chapter. The differential equation

$$x^2 \frac{d^2 y(x)}{dx^2} + (x^2 \cos x + 2x) \frac{dy(x)}{dx} + x \, y(x) = d(x)$$

can be placed in Sturm-Liouville form by multiplying both sides of this equation by a function $f(x)$.

(a) What $f(x)$ will accomplish this?

(b) What is the final Sturm-Liouville form for this equation?

54. If a particular differential equation is not in Sturm-Liouville form (see Exercise 10.53), the Green's function for the equation in the non-Sturm-Liouville form will be different from the Green's function for the equation in Sturm-Liouville

form. Consider the differential equation:

$$x^3 \frac{d^2 y(x)}{dx^2} + x^2 \frac{dy(x)}{dx} - x\, y(x) = d(x),$$

with the boundary conditions $y(0) = y(1) = 0$.

(a) What is the Green's function for this problem as described by the differential equation given above?

(b) Place the differential equation in Sturm-Liouville form.

(c) What is the Green's function for the differential equation in Sturm-Liouville form?

(d) Formulate the response of the system to a drive $d(t)$ using the Green's functions obtained for parts (a) and (c) above and show that these responses are identical.

55. In this chapter, we have seen cases where the Green's function was symmetric with respect to x and ξ. That is to say,

$$g(x|\xi) = g(\xi|x).$$

In this problem you are being asked to show the general conditions that are necessary for this symmetry to hold. Prove that if the differential equation is in Sturm-Liouville form (see Exercise 10.53) and the boundary conditions are homogeneous, the Green's function will always have this symmetry.

To show that this is the case, begin by writing the differential equation that $g(x|\xi)$ satisfies for two different values of ξ: ξ_1 and ξ_2. Multiply the first by $g(x|\xi_2)$ and the second by $g(x|\xi_1)$ and subtract to obtain:

$$g(x|\xi_2)\frac{d}{dx}\left[p(x)\frac{dg(x|\xi_1)}{dx}\right] - g(x|\xi_1)\frac{d}{dx}\left[p(x)\frac{dg(x|\xi_2)}{dx}\right]$$

$$= g(x|\xi_2)\delta(x - \xi_1) - g(x|\xi_1)\delta(x - \xi_2).$$

Show that this expression becomes

$$\frac{d}{dx}\left[g(x|\xi_2)p(x)\frac{dg(x|\xi_1)}{dx}\right] - \frac{d}{dx}\left[g(x|\xi_1)p(x)\frac{dg(x|\xi_2)}{dx}\right]$$

$$= g(x|\xi_2)\delta(x - \xi_1) - g(x|\xi_1)\delta(x - \xi_2).$$

Now if the solution is to be valid over the range $x_1 < x < x_2$, and the boundary conditions are such that the function or its derivative must go to zero at $x = x_1$ and $x = x_2$, show that the above equation can be integrated to give

$$g(\xi_2|\xi_1) = g(\xi_1|\xi_2).$$

56. Using the steps outlined in the previous problem, determine the symmetry properties of the Green's function that satisfies the differential equation

$$x^3 \frac{d^2 y(x)}{dx^2} + x^2 \frac{dy(x)}{dx} - x\,y(x) = d(x),$$

with the boundary conditions $y(0) = y(1) = 0$. Note that this differential equation is not in Sturm-Liouville form.

11

SOLUTIONS TO LAPLACE'S EQUATION

It would be impossible to discuss all the different types of differential equations encountered in physics and engineering problems. In this chapter, we will concentrate on just one, Laplace's equation, because it occurs in so many types of practical problems. In vector notation, valid in any coordinate system, Laplace's equation is

$$\nabla^2 \Phi = 0. \tag{11.1}$$

We will investigate methods of its solution in the three most common coordinate systems: Cartesian, cylindrical, and spherical. The solution methods we present here will also be applicable to many other differential equations.

11.1   CARTESIAN SOLUTIONS

In Cartesian coordinates, Laplace's equation becomes

$$\nabla^2 \Phi(x, y, z) = \left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right] \Phi(x, y, z) = 0. \tag{11.2}$$

A standard approach to solving this equation is to use the method of separation of variables. In this method, we assume $\Phi(x, y, z)$ can be written as the product of three functions, each of which involves only one of the independent variables. That is,

$$\Phi(x, y, z) = X(x)Y(y)Z(z). \tag{11.3}$$

It is important to note that not all the possible solutions of Laplace's equation in Cartesian coordinates have the special separated form of Equation 11.3. As it turns

out, however, all solutions can be written as a linear sum (of potentially infinite number of terms) of these separable solutions. You will see examples of this as the chapter progresses. Using this form for $\Phi(x, y, z)$, Laplace's equation becomes

$$\frac{1}{X(x)} \frac{\partial^2 X(x)}{\partial x^2} + \frac{1}{Y(y)} \frac{\partial^2 Y(y)}{\partial y^2} + \frac{1}{Z(z)} \frac{\partial^2 Z(z)}{\partial z^2} = 0. \tag{11.4}$$

Each of the three terms on the LHS of Equation 11.4 is a function of only one of the independent variables. This means that, for example, the sum of the last two terms

$$\frac{1}{Y(y)} \frac{\partial^2 Y(y)}{\partial y^2} + \frac{1}{Z(z)} \frac{\partial^2 Z(z)}{\partial z^2} \tag{11.5}$$

cannot depend on x. A quick look back at Equation 11.4 shows that this implies the first term, itself, also cannot be a function of x. It obviously is not a function of y or z either, so it must be equal to a constant:

$$\frac{1}{X(x)} \frac{\partial^2 X(x)}{\partial x^2} = c_x. \tag{11.6}$$

Similar arguments can be made to isolate the other terms:

$$\frac{1}{Y(y)} \frac{\partial^2 Y(y)}{\partial y^2} = c_y \tag{11.7}$$

$$\frac{1}{Z(z)} \frac{\partial^2 Z(z)}{\partial z^2} = c_z. \tag{11.8}$$

The three quantities c_x, c_y, and c_z are called separation constants. Notice that they are not completely independent of one another, because Laplace's equation requires

$$c_x + c_y + c_z = 0. \tag{11.9}$$

Consequently, only two can be chosen arbitrarily.

11.1.1 Solutions of the Separated Equations

Each of the three separated equations (Equations 11.6–11.8) is the same linear, second-order differential equation. The solution to Equation 11.6, for a particular, nonzero value of c_x, is simply

$$X(x) = X_+ e^{+\sqrt{c_x} x} + X_- e^{-\sqrt{c_x} x} \qquad c_x \neq 0, \tag{11.10}$$

where X_+ and X_- are arbitrary amplitudes, which are determined only when we apply the boundary conditions. When this general solution is combined with $Y(y)$ and $Z(z)$, these constants combine with the corresponding $Y_\pm$ and $Z_\pm$ constants to form other constants. For this reason, we will ignore these amplitude constants until the final

form of $\Phi(x, y, z)$ is constructed and initially just concentrate on the functional form of the separated solutions. Equation 11.10 will be written, using a shorthand notation, as

$$X(x) \rightarrow \begin{cases} e^{+\sqrt{c_x}\,x} \\ e^{-\sqrt{c_x}\,x} \end{cases} \qquad c_x \neq 0. \tag{11.11}$$

We now turn our attention to the solutions for $X(x)$ with different separation constants. The constant c_x may be negative and real, positive and real, or complex. Because the solution depends on the square root of c_x, it turns out the results of a complex c_x are covered by a combination of solutions generated by real positive c_x and real negative c_x. Thus, we will limit our attention to real values of c_x. There is also the special case of $c_x = 0$ to consider.

In the case with $c_x = 0$, the differential equation becomes

$$\frac{1}{X(x)} \frac{d^2 X(x)}{dx^2} = 0, \tag{11.12}$$

and the general solution has a special form

$$X(x) = Ax + B, \tag{11.13}$$

where A and B are arbitrary amplitude constants. In our shorthand notation, this is written as

$$X(x) \rightarrow \begin{cases} 1 \\ x \end{cases} \qquad c_x = 0. \tag{11.14}$$

This is qualitatively very different than the form of Equation 11.11. Often these special linear solutions can be ignored because of some physical or mathematical reasoning, but not always, and one must be careful not to forget them.

If c_x is real and greater than zero, we can write $\sqrt{c_x} = +\alpha$ where α is real and positive. The solutions for $X(x)$ take the form of growing or decaying exponentials

$$X(x) \rightarrow \begin{cases} e^{+\alpha x} \\ e^{-\alpha x} \end{cases} \qquad c_x = \alpha^2 > 0. \tag{11.15}$$

These solutions are shown in Figure 11.1. Often, it is convenient to rearrange things and use hyperbolic functions instead:

$$X(x) \rightarrow \begin{cases} \sinh(\alpha x) \\ \cosh(\alpha x) \end{cases} \qquad c_x = \alpha^2 > 0. \tag{11.16}$$

The two forms in Equations 11.15 and 11.16 are completely equivalent. Convenience dictates which set is the better choice. Some boundary conditions are more easily accommodated by the exponential solutions, and others by the hyperbolic functions. The sinh function has odd symmetry and is zero at $x = 0$. The cosh function has even symmetry, and is one at $x = 0$. These functions are shown in Figure 11.2. If a
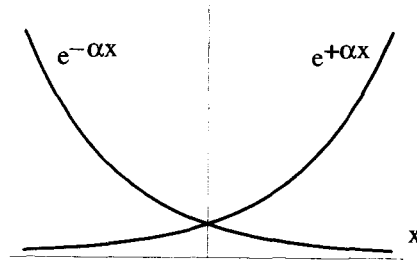
Figure 11.1 Exponential Solutions to Laplace's Equation in Cartesian Geometry

problem naturally has odd or even symmetry, the hyperbolic functions are usually a wise choice.

If c_x is a real negative number, its square root is imaginary. That is, $\sqrt{c_x} = ik$ with $k > 0$. Therefore, for a negative separation constant, the solutions are oscillatory:

$$X(x) \rightarrow \begin{cases} e^{+ikx} \\ e^{-ikx} \end{cases} \qquad c_x = -k^2 < 0. \tag{11.17}$$

For solutions that must be pure real functions, this form is usually clumsy, because the amplitude constants need to be complex. Equation 11.17 can be rewritten using sin and cos functions:

$$X(x) \rightarrow \begin{cases} \sin(kx) \\ \cos(kx) \end{cases} \qquad c_x = -k^2 < 0. \tag{11.18}$$

Again, the choice between the two forms is one of convenience.

In summary, the separated solutions to Laplace's equation in Cartesian geometry have three possible forms. If the separation constant is zero, the solution has the linear form $Ax + B$, if the constant is negative, the solution oscillates, and finally, if the constant is positive, the solution is an exponential. Notice, because the separated differential equations for $Y(y)$ and $Z(z)$ are identical to the equation for $X(x)$, their solutions follow a similar pattern. An interesting condition is imposed on these three solutions by Equation 11.9. If the separated solution in one direction has an
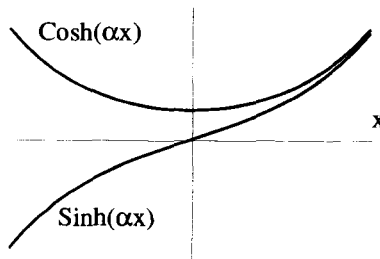
Figure 11.2 The Hyperbolic Solutions to Laplace's Equation in Cartesian Geometry

exponential form, at least one of the other solutions must have an oscillatory form and vice versa. We will discover that this is a general trait of the separated solutions in other coordinate systems as well.

11.1.2   The General Solution

The solution for $\Phi(x, y, x)$ is the product of the three separated solutions, as described by Equation 11.3. The amplitude and separation constants are determined by a set of boundary conditions. These boundary conditions may allow a particular separation constant to take on more than one value and, in many cases, an infinite set of discrete values. In such a situation, because Laplace's equation is linear, the general solution will be composed of a linear sum over the acceptable values of the separation constants.

Let the possible nonzero values of c_x be indexed by m. That is, the mth allowed separation constant for x is c_{xm}. Likewise, let c_{yn} be the nth nonzero separation constant for the y variable. Equation 11.9 implies that c_z cannot be independently specified, because $c_z = -c_{xm} - c_{yn}$. Using the shorthand notation, we can write the general solution as a sum over the two indices n and m:

$$\Phi(x, y, z) = \sum_{n,m} \begin{Bmatrix} e^{+\sqrt{c_{nx}}\,x} \\ e^{-\sqrt{c_{nx}}\,x} \end{Bmatrix} \begin{Bmatrix} e^{+\sqrt{c_{my}}\,y} \\ e^{-\sqrt{c_{my}}\,y} \end{Bmatrix} \begin{Bmatrix} e^{+\sqrt{-c_{nx}-c_{my}}\,z} \\ e^{-\sqrt{-c_{nx}-c_{my}}\,z} \end{Bmatrix} \qquad (11.19)$$

$$+ \text{ The Special Linear Solutions.}$$

The sum on the RHS of this equation is a shorthand notation for writing all the combinations of solutions that have nonzero separation constants, each with an un-determined amplitude constant. For each value of m and n, there are eight possible combinations of the solution types. If you wanted to expand this out, it would look like this:

$$\sum_m \sum_n \begin{Bmatrix} A_{mn}e^{+\sqrt{c_{nx}}\,x}e^{+\sqrt{c_{my}}\,y}e^{+\sqrt{-c_{nx}-c_{my}}\,z} \\ +B_{mn}e^{-\sqrt{c_{nx}}\,x}e^{+\sqrt{c_{my}}\,y}e^{+\sqrt{-c_{nx}-c_{my}}\,z} \\ +C_{mn}e^{+\sqrt{c_{nx}}\,x}e^{-\sqrt{c_{my}}\,y}e^{+\sqrt{-c_{nx}-c_{my}}\,z} \\ +D_{mn}e^{-\sqrt{c_{nx}}\,x}e^{-\sqrt{c_{my}}\,y}e^{+\sqrt{-c_{nx}-c_{my}}\,z} \\ +E_{mn}e^{+\sqrt{c_{nx}}\,x}e^{+\sqrt{c_{my}}\,y}e^{-\sqrt{-c_{nx}-c_{my}}\,z} \\ +F_{mn}e^{-\sqrt{c_{nx}}\,x}e^{+\sqrt{c_{my}}\,y}e^{-\sqrt{-c_{nx}-c_{my}}\,z} \\ +G_{mn}e^{+\sqrt{c_{nx}}\,x}e^{-\sqrt{c_{my}}\,y}e^{-\sqrt{-c_{nx}-c_{my}}\,z} \\ +H_{mn}e^{-\sqrt{c_{nx}}\,x}e^{-\sqrt{c_{my}}\,y}e^{-\sqrt{-c_{nx}-c_{my}}\,z} \end{Bmatrix}. \qquad (11.20)$$

Many times the symmetry of the problem will eliminate some of the terms in the bracketed sum of Equation 11.20, and those particular amplitude constants will be zero. For this reason, it is convenient to use the shorthand notation of Equation 11.19 and not address the amplitude constants until the symmetry arguments have placed the general solution in more tractable form.

The "special" solutions of Equation 11.19 are the ones that fall outside of normal form because one or more of the separation constants is zero. For example, if a possible solution has $c_x = 0$, but $c_y = c$ and $c_z = -c$, there is a special set of linear solutions

$$\begin{Bmatrix} 1 \\ x \end{Bmatrix} \begin{Bmatrix} e^{+\sqrt{c}y} \\ e^{-\sqrt{c}y} \end{Bmatrix} \begin{Bmatrix} e^{+\sqrt{-c}z} \\ e^{-\sqrt{-c}z} \end{Bmatrix} . \qquad (11.21)$$

When $c_x = c_y = c_z = 0$, another set of special solutions is

$$\begin{Bmatrix} 1 \\ x \end{Bmatrix} \begin{Bmatrix} 1 \\ y \end{Bmatrix} \begin{Bmatrix} 1 \\ z \end{Bmatrix} . \qquad (11.22)$$

The general solution will be unique, that is to say all the amplitude and separation constants will be determined, only if enough boundary conditions are specified. For Laplace's equation this means that the solution will be unique in a region if the value of $\Phi(x, y, z)$ or its normal derivative, or a combination of $\Phi(x, y, z)$ and its normal derivative is specified over the entire surface enclosing that region. These are referred to as the Neumann or Dirichlet boundary conditions, respectively.

Example 11.1 To see how the boundary conditions are used to determine the constants of Equation 11.19, consider the two-dimensional problem depicted in Figure 11.3. Two parallel plates exist in the planes $y = +h$ and $y = -h$, with periodic square wave electric potentials applied as shown. The problem is two dimensional, because there are no variations in the z-direction. The potential along the x-direction repeats with a spatial period of $2d$. The boundary conditions over the first period can be summarized as

$$\Phi(x, +h) = \begin{cases} +1 & 0 < x < d \\ -1 & d < x < 2d \end{cases} \qquad (11.23)$$
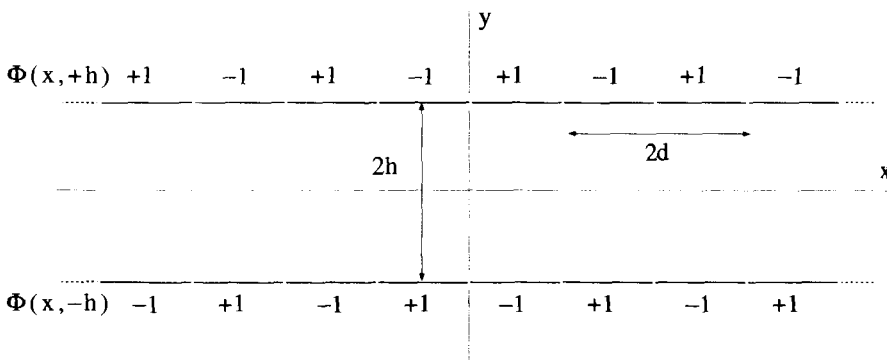
Figure 11.3 Two-Dimensional Laplace Equation Problem with Square Wave Boundary Conditions

and

$$\Phi(x, -h) = \begin{cases} -1 & 0 < x < d \\ +1 & d < x < 2d \end{cases}. \tag{11.24}$$

We want to find the solution for the potential between the plates, using the boundary conditions and the two-dimensional form of Laplace's equation:

$$\frac{\partial^2 \Phi(x, y)}{\partial x^2} + \frac{\partial^2 \Phi(x, y)}{\partial y^2} = 0. \tag{11.25}$$

Based on our previous discussion, we will look for a two-dimensional solution with the form

$$\Phi(x, y) = X(x)Y(y). \tag{11.26}$$

A quick look at the nature of the boundary conditions shows that the solution of $X(x)$ must be oscillatory in order to satisfy the periodic nature of the problem. Therefore, it must use a negative, real separation constant:

$$\frac{1}{X(x)} \frac{d^2 X(x)}{dx^2} = -k^2, \tag{11.27}$$

where we have put the constant in the form of a square to make the subsequent algebra simpler. The other solution must therefore obey

$$\frac{1}{Y(y)} \frac{d^2 Y(y)}{dy^2} = +k^2. \tag{11.28}$$

The possible solutions to Equation 11.27 are constrained by two facts. First, because the boundary conditions of the problem have odd symmetry about x, $X(x)$ must also have odd symmetry. Also, because the boundary conditions of the problem have a periodicity of $2d$, each solution of $X(x)$ must oscillate with a period of $2d/n$ where n is any positive integer. This ensures $X(x + 2d) = X(x)$. All the possible solutions of $X(x)$ can be written using n as an index,

$$X(x) \rightarrow \sin(k_n x), \tag{11.29}$$

where $k_n = \pi n/d$. Notice the $k = 0$ solution is not allowed, since the corresponding solution of $X(x) = Ax + B$ would violate the combination of the symmetry and periodic requirements. This means we do not need to consider the special linear solutions we mentioned in the previous section. For the $Y(y)$ solution, we will use the sinh function, because it matches the odd symmetry of the boundary conditions in the y-direction:

$$Y(y) \rightarrow \sinh(k_n y). \tag{11.30}$$

The combination of these two solutions generates the general solution for $\Phi(x, y)$,

$$\Phi(x, y) = \sum_{n=1,2,\ldots}^{\infty} A_n \sin(k_n x) \sinh(k_n y), \tag{11.31}$$

where the amplitude constants A_n are finally introduced.

The A_n coefficients are determined by requiring Equation 11.31 go to the applied potential on the boundaries. Since we have already imposed the odd symmetry in the y-direction, if the expression in Equation 11.31 goes to the proper potential at $y = +h$, it will automatically go to the correct value at $y = -h$ as well. Thus to make everything work, the solution must satisfy

$$\sum_{n=1,2,\ldots}^{\infty} A_n \sin\left(\frac{\pi n x}{d}\right) \sinh\left(\frac{\pi n h}{d}\right) = \begin{cases} +1 & 0 < x < d \\ -1 & d < x < 2d \end{cases}. \tag{11.32}$$

The LHS of Equation 11.32 is in the form of a Fourier sine series. A particular coefficient A_m can be obtained by operating on both sides of this equation with

$$\int_0^{2d} dx \, \sin\left(\frac{\pi m x}{d}\right). \tag{11.33}$$

Every term in the series drops out except the $n = m$ term, giving the result:

$$A_m = \begin{cases} 0 & m \text{ even} \\ \dfrac{4}{m\pi \, \sinh(m\pi h/d)} & m \text{ odd} \end{cases}. \tag{11.34}$$

The complete solution to the problem is then

$$\Phi(x, y) = \sum_{n=1,3,\ldots}^{\infty} \frac{4}{n\pi \, \sinh(n\pi h/d)} \sin\left(\frac{n\pi x}{d}\right) \sinh\left(\frac{n\pi y}{d}\right). \tag{11.35}$$

A plot of the first 21 terms of the RHS of Equation 11.35, for the region $-2d < x < 2d$ and $-h < y < h$, is shown in Figure 11.4. Notice the Gibbs overshoot phenomenon occurring at the discontinuities.

Example 11.2 The previous example was particularly simple, because the periodic boundary conditions immediately forced one of the separated solutions to have an oscillatory form. The boundary conditions shown in Figure 11.5 are not as obliging. This is again a two-dimensional problem, but now we are looking for the solution to Laplace's equation in the interior of an $a \times b$ box. On the two vertical surfaces, the potential is required to be zero, while on the top and bottom surfaces, $\Phi = \pm 1$. In this case, it isn't obvious that either $X(x)$ or $Y(y)$ needs to have an oscillatory form, because none of the boundary conditions are periodic. We can use a common trick, however, and attempt to convert this problem into an equivalent periodic problem.
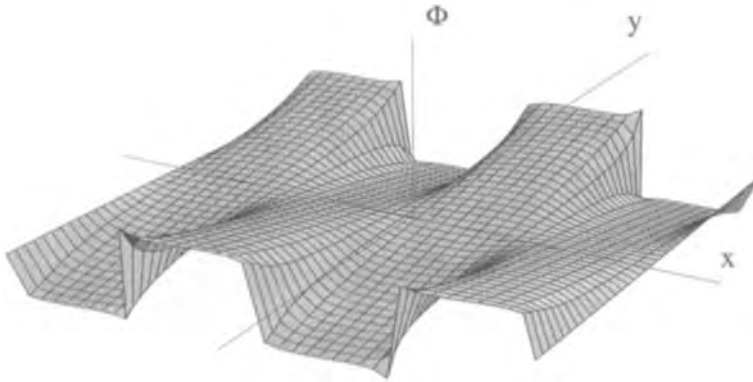
Figure 11.4   The Electric Potential Distribution for Square Wave Boundary Conditions at $y = \pm 1$
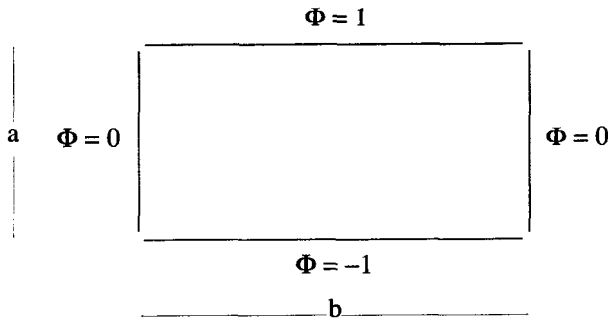

Figure 11.5   Square Box Boundary Conditions

We will try to change the problem into the form shown in Figure 11.6. The position of the original $a \times b$ box is indicated by the shaded region shown. The top and bottom boundaries of this shaded region must be fixed at $\Phi = \pm 1$ so they match the original boundary conditions. We need to apply periodic conditions along the rest of the top and bottom surfaces, where the ? marks appear in the figure, to try to force the vertical
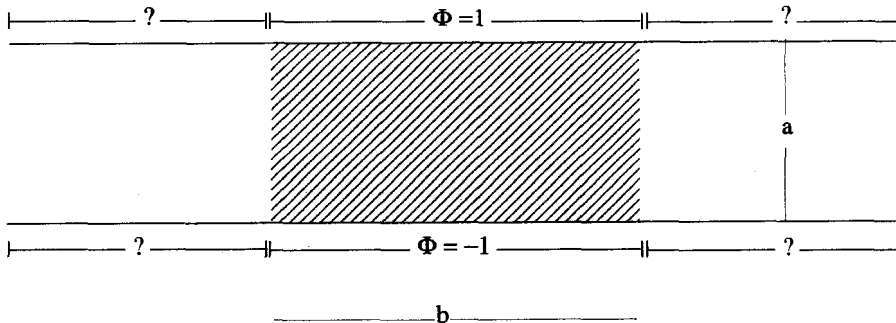
Figure 11.6   Attempt to Convert to a Problem Periodic in the x-Direction

$\Phi = 1$	$\Phi = -1$	$\Phi = 1$	$\Phi = -1$	$\Phi = 1$

a

$\Phi = -1$	$\Phi = 1$	$\Phi = -1$	$\Phi = 1$	$\Phi = -1$

— · — b — · — ·
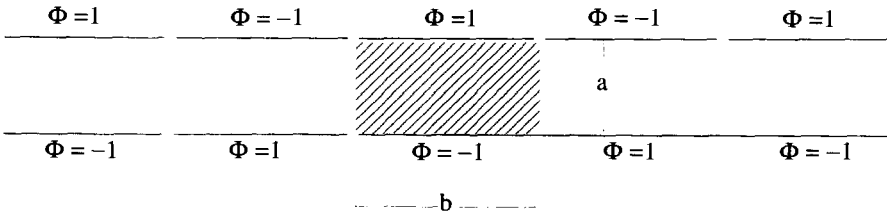
Figure 11.7   Conversion to a Periodic Problem in the x-Direction

sides of the shaded region to have $\Phi = 0$. Looking back on the previous example, it is pretty easy to see that these requirements are met with the potential distribution shown in Figure 11.7. The solution inside the box is therefore identical to the solution to the previous example with the x-periodicity equal to $2b$. The only difference is, of course, that this solution is only valid inside the region of the original $a \times b$ box of Figure 11.5.

We successfully made this problem periodic in the x-direction. Could we have done the same thing in the y-direction? Figure 11.8 shows what this would look like. Again, we need to apply periodic potentials to the areas where the ? marks occur in the figure, to try to force the horizontal sides of the shaded region to have $\Phi = \pm 1$. It's pretty obvious this cannot work. In this case, the x-direction must be chosen as the periodic one.

An interesting twist is added to this problem if we modify the boundary conditions, as shown in Figure 11.9. In a case like this, you will find it is impossible to simply make either the x- or y-direction periodic. Instead, the problem must be broken up into two parts, as shown in Figure 11.10. Because Laplace's Equation is linear, the solution can be written as the sum of two solutions, $\Phi_1(x, y)$ and $\Phi_2(x, y)$, which each satisfy modified boundary conditions shown for the two boxes on the right. The function $\Phi_1(x, y)$ satisfies Laplace's equation and has boundary conditions which are amenable to periodic extension in the x-direction. The function $\Phi_2(x, y)$ also satisfies Laplace's equation, but with conditions which can be extended in the y-direction. The combination $\Phi_1(x, y) + \Phi_2(x, y) = \Phi(x, y)$ satisfies Laplace's equation inside the box and satisfies the original boundary conditions of Figure 11.9.

11.2   EXPANSIONS WITH EIGENFUNCTIONS

Before looking at solutions to Laplace's equation in other coordinate systems, we need to establish some general techniques for expanding functions with linear summations of orthogonal functions. In the Cartesian problems we have discussed so far, we found that we could evaluate the individual amplitude constants in the general solution by using Fourier series type orthogonality conditions. This worked because the terms in our sums of the general solution contained sine and cosine functions. In this section, we will generalize this technique for other problems. We will begin by reviewing the
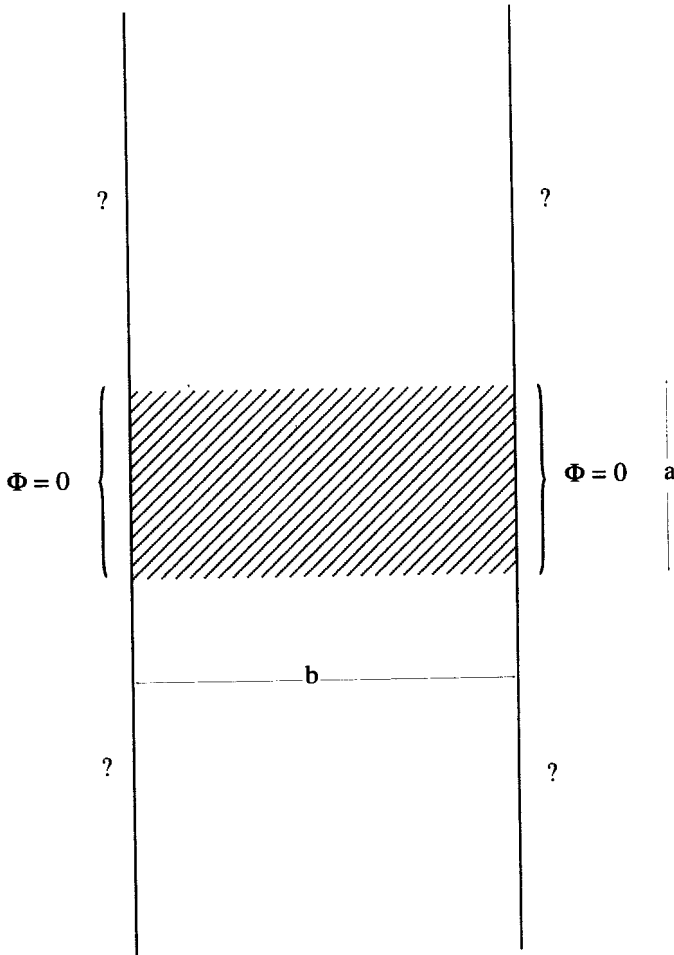
Figure 11.8   Improper Conversion to a Periodic Problem in the y-Direction
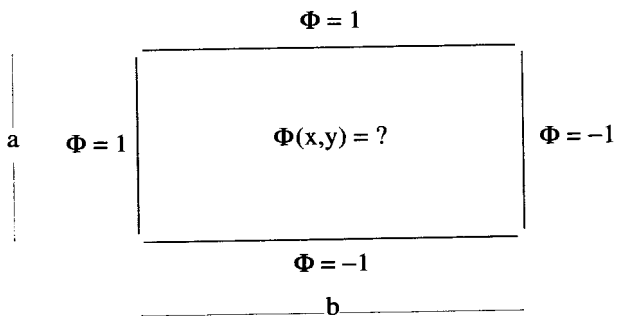

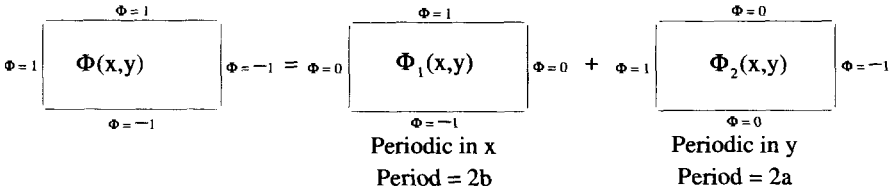**Figure 11.9**   Compound Cartesian Boundary Conditions

Figure 11.10 Dividing the Compound Problem into Two Periodic Problems

eigenvalue and eigenvector concepts associated with the linear algebra of matrices, with which the reader is assumed to be familiar. These ideas will then be extended to the eigenvalues and eigenfunctions of linear differential operators.

11.2.1   Eigenvectors and Eigenvalues

Eigenvectors and eigenvalues are concepts used frequently in matrix algebra. A matrix $[M]$ is said to be associated with a set of eigenvectors $[V_n]$ and eigenvalues λ_n, if

$$[M][V_n] = \lambda_n [V_n]. \tag{11.36}$$

In other words, a vector is an eigenvector of a matrix if premultiplying it by the matrix results in a constant times the original vector. The proportionality constant is the eigenvalue. Given the matrix $[M]$, the eigenvalues are found from the condition

$$|[M] - [1]\lambda_n|_{det} = 0. \tag{11.37}$$

The eigenvectors, to within an arbitrary constant, can then be found by inserting the eigenvalues into Equation 11.36 and solving for the components. A 3×3 matrix will have three eigenvalues. If the eigenvalues are all different, there will be three distinct eigenvectors. If there is some degeneracy in the eigenvalues, then the eigenvectors associated with these degenerate eigenvalues will not be uniquely defined. These ideas were covered in detail in Chapter 4.

A matrix $[M]$ is Hermitian if

$$\underline{M}_{ij} = \underline{M}_{ji}^*. \tag{11.38}$$

That is to say, a matrix is Hermitian if it is equal to its complex conjugate transpose. If a matrix is Hermitian, its eigenvalues are always pure real, and its eigenvectors are orthogonal.

11.2.2   Eigenfunctions and Eigenvalues

These ideas can be extended to operators. Let $\mathcal{L}_{op}(x)$ be a linear differential operator. In direct analogy with the matrix definition, the set of eigenfunctions $\phi_n(x)$ associated with this operator satisfy the condition

$$\mathcal{L}_{op}(x)\,\phi_n(x) = \lambda_n \phi_n(x). \tag{11.39}$$

In other words, when the operator acts on any of its eigenfunctions, the result is just a constant times the eigenfunction. The constant λ_n is still referred to as an eigenvalue.

Actually, it is convenient to extend this definition a little further, to include what is called a weighting function. The eigenfunctions of an operator with a weighting function obey the equation

$$\mathcal{L}_{op}(x)\ \phi_n(x) = \lambda_n w(x)\phi_n(x), \tag{11.40}$$

where $w(x)$ is any real, *positive definite* function of x. Positive definite is a fancy way of saying $w(x) > 0$ for all x. Our original definition in Equation 11.39 is a just special case where $w(x) = 1$.

11.2.3   Hermitian Operators

An operator $\mathcal{L}_{op}(x)$ can be combined with a set of basis functions $\underline{u}_n(x)$ to form a set of matrix elements

$$\underline{L}_{ij} = \int_{\Omega} dx\ \underline{u}_i^*(x)\mathcal{L}_{op}\underline{u}_j(x). \tag{11.41}$$

The functions $\underline{u}_n(x)$ may or may not be the eigenfunctions of the operator. In general, these functions are complex, as is the operator itself. The matrix elements generated by this integration can also be complex. The integration is done over some range of the variable x that we label with the symbol Ω.

Following an analogy with matrix algebra, a linear operator is Hermitian if

$$\underline{L}_{ij} = \underline{L}_{ji}^*, \tag{11.42}$$

or equivalently

$$\int_{\Omega} dx\ \underline{u}_i^*(x)\mathcal{L}_{op}\underline{u}_j(x) = \int_{\Omega} dx\ \underline{u}_j(x)\left[\mathcal{L}_{op}\underline{u}_i(x)\right]^*. \tag{11.43}$$

In general, satisfying the Hermitian condition of Equation 11.43 depends not only on the operator $\mathcal{L}_{op}$, but also on the functions $\underline{u}_n(x)$ and specifically the nature of their boundary conditions at the limits of Ω. Most often, it is convenient to use the eigenfunctions of the operator for the $\underline{u}_n(x)$, but in some cases other functions are used.

11.2.4   Eigenvalues and Eigenfunctions of Hermitian Operators

The eigenvalues and eigenfunctions of a Hermitian operator have three very useful properties. First the eigenvalues are always pure real. Second, the eigenfunctions will always obey an orthogonality condition. Third, the collection of eigenfunctions usually forms a complete set of functions. That is, any function can be expanded as a linear sum of the eigenfunctions.

To prove the first two of these assertions, let $\phi_n(x)$ be the nth eigenfunction of the Hermitian operator $\mathcal{L}_{op}$, with a corresponding eigenvalue of λ_n. We will use the

extended version of the eigenvalue/eigenfunction condition that includes a weighting function:

$$\mathcal{L}_{op}\phi_n(x) = \lambda_n w(x)\phi_n(x). \tag{11.44}$$

Remember this weighting function is a real, positive definite function. If the Hermitian condition of Equation 11.43 is satisfied, then Equation 11.44 can be substituted into this condition to give

$$\int_{\Omega} dx \ \phi_m^*(x)\lambda_n w(x)\phi_n(x) = \int_{\Omega} dx \ \phi_n(x)\lambda_m^* w(x)\phi_m^*(x), \tag{11.45}$$

which, on rearrangement, becomes

$$(\lambda_n - \lambda_m^*) \int_{\Omega} dx \ \phi_m^*(x)w(x)\phi_n(x) = 0. \tag{11.46}$$

This equation contains some very useful information. If $m = n$, the quantity $\phi_n^*(x)w(x)\phi_n(x) = |\phi_n(x)|^2 w(x)$ is positive definite, so

$$\int_{\Omega} dx \ \phi_n^*(x)w(x)\phi_n(x) \neq 0. \tag{11.47}$$

Consequently,

$$\lambda_n = \lambda_n^*. \tag{11.48}$$

This shows the eigenvalues of a Hermitian operator are real. In addition, if $m \neq n$ and $\lambda_n \neq \lambda_m$, the only way Equation 11.46 can work is if

$$\int_{\Omega} dx \ \phi_m^*(x)w(x)\phi_n(x) = 0. \tag{11.49}$$

Equation 11.49 is an orthogonality condition. Notice the range of the integral Ω is the same range used in Equation 11.43 to produce the Hermitian condition.

If two or more eigenfunctions have the same eigenvalue, that is, if $\lambda_n = \lambda_m$, the line of reasoning which leads to Equation 11.49 is not valid. We can, however, always arrange our choices for the degenerate eigenfunctions so that Equation 11.49 still holds. This is directly analogous to the case of degenerate eigenvalues in matrix algebra, discussed in Chapter 4. The general method for constructing linearly independent, orthogonal eigenfunctions that have degenerate eigenvalues is called the Gram-Schmidt orthogonalization process. You can work out an example of this process in one of the exercises at the end of the chapter.

We also claimed that the eigenfunctions of a Hermitian operator will usually form the basis of a complete set of functions. Completeness is a difficult property to prove and beyond the scope of this text. It can be shown that the eigenfunctions of a Sturm-Liouville operator always form a complete set. These types of operators are discussed in the following section.

The consequences of having a complete set of orthogonal eigenfunctions are enormous. If this is the case, we have a method for expanding arbitrary functions in terms of these eigenfunctions. Imagine we wanted to expand the function $s(x)$ using the eigenfunctions of an operator that obeys Equation 11.44. If the eigenfunctions are complete, we can write

$$s(x) = \sum_n \underline{S}_n \phi_n(x), \tag{11.50}$$

where the $\underline{S}_n$ are (possibly complex) coefficients that we need to determine. To isolate the mth coefficient, multiply both sides of this expression by $w(x)\phi_m^*(x)$ and integrate over the full range:

$$\int_\Omega dx\, s(x)w(x)\phi_m^*(x) = \int_\Omega dx\, \left[\sum_n \underline{S}_n\phi_n(x)\right] w(x)\phi_m^*(x). \tag{11.51}$$

The orthogonality condition will force every element of the sum on the right-hand side to vanish, except the $n = m$ term. That single remaining term gives an expression for $\underline{S}_m$:

$$\underline{S}_m = \frac{\int_\Omega dx\, s(x)w(x)\phi_m^*(x)}{\int_\Omega dx\, \phi_m(x)w(x)\phi_m^*(x)}. \tag{11.52}$$

11.2.5   Sturm-Liouville Operators

We will now prove that the Hermitian condition always holds for the so-called Sturm-Liouville operators, if a special set of boundary conditions is imposed on the functions used in the Hermitian integration. An operator is in Sturm-Liouville form if it can be written as

$$\mathcal{L}_{op} = \frac{d}{dx}\left[p(x)\frac{d}{dx}\right] + q(x), \tag{11.53}$$

where $p(x)$ and $q(x)$ are two real functions. This is not a very restrictive condition, because the differential operations in all second-order linear equations can be converted into this form by multiplying the differential equation by the proper function.

To see how the Sturm-Liouville form of Equation 11.53 leads to the Hermitian condition, substitute this operator into the LHS of Equation 11.43 to obtain

$$\int_\Omega dx\, u_i^*(x)\frac{d}{dx}\left[p(x)\frac{d}{dx}u_j(x)\right] + \int_\Omega dx\, u_i^*(x)q(x)u_j(x). \tag{11.54}$$

Likewise, the RHS of Equation 11.43 becomes

$$\int_\Omega dx\, u_j(x)\frac{d}{dx}\left[p(x)\frac{d}{dx}u_i^*(x)\right] + \int_\Omega dx\, u_j(x)q(x)u_i^*(x). \tag{11.55}$$

The last terms of both these expressions are obviously the same, so to prove Equation 11.43, we only need show the equality of the first two terms. Integrating these terms by parts and equating them gives

$$\left[u_i^*(x)p(x)\frac{du_j(x)}{dx} \right] \Bigg|_\Omega - \int_\Omega dx \, \frac{du_i^*(x)}{dx}p(x)\frac{du_j(x)}{dx}$$

$$= \left[u_j(x)p(x)\frac{du_i^*(x)}{dx} \right] \Bigg|_\Omega - \int_\Omega dx \, \frac{du_j(x)}{dx}p(x)\frac{du_i^*(x)}{dx}. \tag{11.56}$$

The integrals in Equation 11.56 cancel, so if we limit our possible choice of functions to those that satisfy

$$\left[u_n(x)p(x)\frac{du_m^*(x)}{dx} \right] \Bigg|_\Omega = 0 \tag{11.57}$$

for all possible choices of $u_m(x)$ and $u_n(x)$, the Hermitian condition will always be satisfied.

It is important to note that an operator that is not in Sturm-Liouville form can still be Hermitian. A good example of this is the operator $\mathcal{L} = -i\partial/\partial x$, which you can easily show is Hermitian when appropriate boundary conditions are imposed on its eigenfunctions. We presented the proof for Sturm-Liouville systems here for several reasons. First, all the second-order, separated differential equations that are used to solve Laplace's equation can be put into Sturm-Liouville form. In addition, although we will not prove it here, if an operator is in Sturm-Liouville form, its eigenfunctions form a complete set of functions. The proof of this fact can be found in Griffel's book, *Applied Functional Analysis*. Consequently, a linear sum of these complete, orthogonal solutions can be used to satisfy any set of well posed boundary conditions.

Example 11.3   Consider the simple differential operator

$$\mathcal{L}_{op} = \frac{d^2}{dx^2}. \tag{11.58}$$

For the simplest weighting function $w(x) = 1$, the eigenfunctions and eigenvalues of this operator satisfy the equation

$$\frac{d^2}{dx^2}\,\phi_n(x) = \lambda_n\phi_n(x). \tag{11.59}$$

We already encountered this equation when we separated Laplace's equation in Cartesian coordinates. Recall, there were several possible forms for the general solution of Equation 11.59. One form was

$$\phi_n(x) = \underline{A}_n e^{+\sqrt{\lambda_n}x} + \underline{B}_n e^{-\sqrt{\lambda_n}x}, \tag{11.60}$$

where $\underline{A}_n$ and $\underline{B}_n$ are complex amplitude constants. Since we have not yet imposed any other requirements beyond the differential equation, the eigenvalues and amplitude constants are undetermined.

What happens if we check to see if $\mathcal{L}_{op}$ is Hermitian? First, notice that this operator is in the Sturm-Liouville form of Equation 11.53, with $p(x) = 1$ and $q(x) = 0$. This means the Hermitian condition is satisfied, using the eigenfunctions as the basis function, as long as the $\phi_n(x)$ satisfy the condition of Equation 11.57:

$$\phi_n \frac{d\phi_m^*}{dx}\bigg|_a^b = 0. \tag{11.61}$$

Notice that the range of integration Ω has been defined by the two end points $x = a$ and $x = b$. The easiest way to satisfy this expression is to require the $\phi_n(x)$ or their first derivatives be zero at these endpoints. The only way to accomplish this is to make $\phi_n(x)$ oscillatory, so the λ_n must be real, negative numbers. If we define $ik_n \equiv \sqrt{\lambda_n}$, this puts Equation 11.60 into the form

$$\phi_n(x) = \underline{A}_n e^{ik_n x} + \underline{B}_n e^{-ik_n x}. \tag{11.62}$$

If we specifically force $\phi_n(a) = 0$ and $\phi_n(b) = 0$, the eigenvalues are limited to the discrete set of values

$$\lambda_n = -\left(\frac{\pi n}{b-a}\right)^2 \quad n = 1, 2, 3 \ldots, \tag{11.63}$$

and the most succinct form for the eigenfunctions becomes

$$\phi_n(x) = \underline{C}_n \sin\left(\frac{n\pi(x-a)}{b-a}\right). \tag{11.64}$$

The amplitude constant $\underline{C}_n$ is arbitrary, but is usually chosen so the eigenfunction is normalized over the range of Ω. In this particular case, we can set

$$\phi_n(x) = \sqrt{\frac{2}{b-a}} \sin\left(\frac{n\pi(x-a)}{b-a}\right). \tag{11.65}$$

According to our previous discussion, these eigenfunction should obey an orthogonality condition. Indeed this is the case, because it is easy to show that

$$\int_a^b dx \, \phi_n(x)\phi_m(x) = \delta_{nm}. \tag{11.66}$$

11.3 CYLINDRICAL SOLUTIONS

Now that we know how to perform expansions using the eigenfunctions of Hermitian operators, we can proceed to solve Laplace's equation in other coordinate systems. In this section, we explore the solutions to Laplace's equation in cylindrical geometries. In cylindrical coordinates Laplace's equation becomes

$$\nabla^2 \Phi(\rho, \theta, z) = \left[\frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2} \right] \Phi(\rho, \theta, z) = 0. \qquad (11.67)$$

We will again use the method of separation of variables and look for solutions of the form

$$\Phi(\rho, \theta, z) = R(\rho)H(\theta)Z(z). \qquad (11.68)$$

Inserting this form into Laplace's equation gives

$$\frac{1}{R(\rho)} \frac{\partial^2 R(\rho)}{\partial \rho^2} + \frac{1}{\rho R(\rho)} \frac{\partial R(\rho)}{\partial \rho} + \frac{1}{\rho^2 H(\theta)} \frac{\partial^2 H(\theta)}{\partial \theta^2} + \frac{1}{Z(z)} \frac{\partial^2 Z(z)}{\partial z^2} = 0. \qquad (11.69)$$

11.3.1 Solutions of the Separated Equations

In cylindrical coordinates, things are quite a bit more complicated than they were in the Cartesian case, because the ρ and θ dependencies are not isolated in Equation 11.69. In addition, unlike the Cartesian solutions, the ρ, θ, and z separated solutions do not all have the same functional form. We will tackle these different solutions one at a time.

Separating the z-Dependence The z-dependence is easy to separate because it is entirely contained in the last term of Equation 11.69. Setting this term equal to a separation constant c_z gives

$$\frac{1}{Z(z)} \frac{\partial^2 Z(z)}{\partial z^2} = c_z. \qquad (11.70)$$

This has the form of an eigenvalue problem

$$\frac{\partial^2 Z(z)}{\partial z^2} = c_z Z(z). \qquad (11.71)$$

Just like the Cartesian examples we presented earlier in this chapter, the possible values that c_z can take on will be limited by whatever boundary conditions are specified for the problem. For example, perhaps $\Phi(\rho, \theta, z)$ is required to be zero at $z = a$ and $z = b$. We explored this case already in a previous example. This forces $Z(z)$ to have an oscillatory form, and consequently c_z must be negative and real. If

we set $c_z = -k^2 < 0$, the form of $Z(z)$ is

$$Z(z) \rightarrow \begin{cases} e^{+ikz} \\ e^{-ikz} \end{cases} \quad \text{or} \quad Z(z) \rightarrow \begin{cases} \sin(kz) \\ \cos(kz) \end{cases} \quad c_z = -k^2 < 0. \qquad (11.72)$$

We have already shown that this operator, when combined with the functions of Equation 11.72 and the proper boundary conditions, is Hermitian. Therefore, we can expand arbitrary functions of z using this set of solutions.

In contrast, we might have a set of boundary conditions which requires $\Phi(\rho, \theta, z)$ to vanish as $z \rightarrow \pm\infty$. Oscillatory solutions of $Z(z)$ are obviously not going to work in this case. In fact, the only way to get a solution that meets these conditions is to use a positive, real separation constant so the solutions of $Z(z)$ are exponential. With $c_z = \alpha^2 > 0$,

$$Z(z) \rightarrow \begin{cases} e^{+\alpha z} \\ e^{-\alpha z} \end{cases} \quad \text{or} \quad Z(z) \rightarrow \begin{cases} \sinh(\alpha z) \\ \cosh(\alpha z) \end{cases} \quad c_z = \alpha^2 > 0. \qquad (11.73)$$

These solutions, when combined with the $\partial^2/\partial z^2$ operator, do not satisfy the Hermitian condition and consequently do not obey an orthogonality condition. This means it is not possible to use these functions in an expansion.

As always, the special case where $c_z = 0$ needs to be considered. The form of $Z(z)$ is linear:

$$Z(z) \rightarrow \begin{cases} 1 \\ z \end{cases} \quad c_z = 0. \qquad (11.74)$$

Separating the θ-Dependence   After removing the z-dependence and multiplying by ρ^2, Equation 11.69 becomes

$$\frac{\rho^2}{R(\rho)} \frac{\partial^2 R(\rho)}{\partial \rho^2} + \frac{\rho}{R(\rho)} \frac{\partial R(\rho)}{\partial \rho} + \rho^2 c_z + \frac{1}{H(\theta)} \frac{\partial^2 H(\theta)}{\partial \theta^2} = 0. \qquad (11.75)$$

The θ-dependence is now isolated to the last term, which can be set equal to a second separation constant c_θ:

$$\frac{1}{H(\theta)} \frac{\partial^2 H(\theta)}{\partial \theta^2} = c_\theta. \qquad (11.76)$$

This equation is also in the form of an eigenvalue problem, identical to the one describing the z-dependence. Taking $c_\theta = -\nu^2 < 0$ gives orthogonal, oscillating solutions of the form

$$H(\theta) \rightarrow \begin{cases} e^{+i\nu\theta} \\ e^{-i\nu\theta} \end{cases} \quad \text{or} \quad H(\theta) \rightarrow \begin{cases} \sin(\nu\theta) \\ \cos(\nu\theta) \end{cases} \quad c_\theta = -\nu^2 < 0. \qquad (11.77)$$

Taking $c_\theta = \beta^2 > 0$ results in nonorthogonal solutions of the form

$$H(\theta) \rightarrow \begin{cases} e^{+\beta\theta} \\ e^{-\beta\theta} \end{cases} \quad \text{or} \quad H(\theta) \rightarrow \begin{cases} \sinh(\beta\theta) \\ \cosh(\beta\theta) \end{cases} \quad c_\theta = \beta^2 > 0. \quad (11.78)$$

The special case of $c_\theta = 0$ gives the form

$$H(\theta) \rightarrow \begin{cases} 1 \\ \theta \end{cases} \quad c_\theta = 0. \quad (11.79)$$

Most cylindrical solutions need to be periodic in θ, repeating every 2π. If this were not the case, $H(\theta) \neq H(\theta + 2\pi)$, and $\Phi(\rho, \theta, z)$ would not be a single-valued function of position. For solutions that are periodic in θ, the form given in Equation 11.77 with $\nu = 1, 2, 3, \ldots$ and the $H(\theta) \rightarrow 1$ solution of Equation 11.79 are the only ones allowed. We can actually combine these two forms together into just the form of Equation 11.77 if we allow $\nu = 0$ to be included in that expression, so now $c_\theta = -\nu^2 \leq 0$. There are, however, some cases where the boundary conditions limit the range of θ to something less than 0 to 2π. In these cases, c_θ can be greater than zero, and the other solutions of Equation 11.78 plus the $H(\theta) \rightarrow \theta$ solution of Equation 11.79 are allowed.

Separating the ρ-Dependence With the z- and θ-dependencies removed by inserting their separation constants, the differential equation for $R(\rho)$ becomes

$$\frac{\rho^2}{R(\rho)} \frac{\partial^2 R(\rho)}{\partial \rho^2} + \frac{\rho}{R(\rho)} \frac{\partial R(\rho)}{\partial \rho} + c_z \rho^2 + c_\theta = 0. \quad (11.80)$$

The form of the solution to this equation depends on the nature of the separation constants. When $c_\theta \leq 0$, the most common case, the solutions of $R(\rho)$ are either Bessel functions or modified Bessel functions depending on the sign of the c_z constant. The sections that follow derive series representations for these important functions.

11.3.2 Solutions for $R(\rho)$ with $c_z > 0$ and $c_\theta \leq 0$

The first situation we will consider is the case for $c_\theta = -\nu^2 \leq 0$ and $c_z = \alpha^2 > 0$. Notice we have included $\nu = 0$ in this category as we discussed when separating out the θ-dependence in the previous section. This is a solution that oscillates (except when $\nu = 0$) in θ and exponentially grows or decays in z. For this case, Equation 11.80 becomes

$$\frac{\partial^2 R(\rho)}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial R(\rho)}{\partial \rho} + \alpha^2 R(\rho) - \frac{\nu^2}{\rho^2} R(\rho) = 0. \quad (11.81)$$

The procedure for solving this equation first involves scaling the ρ variable using the value of α. Notice that these two quantities have reciprocal dimensions so that the quantity $(\alpha\rho)$ is dimensionless. To accomplish this, rearrange Equation 11.81 into the form

$$\frac{\partial^2 R(\rho)}{\partial (\alpha\rho)^2} + \frac{1}{\alpha\rho} \frac{\partial R(\rho)}{\partial (\alpha\rho)} + R(\rho) - \frac{\nu^2}{(\alpha\rho)^2} R(\rho) = 0. \quad (11.82)$$

Now define a new variable $r \equiv \alpha\rho$ and a new function

$$R(\rho) = \grave{R}(\alpha\rho). \tag{11.83}$$

With these new quantities, Equation 11.82 can be written as

$$\frac{\partial^2 \grave{R}(r)}{\partial r^2} + \frac{1}{r}\frac{\partial \grave{R}(r)}{\partial r} + \grave{R}(r) - \frac{\nu^2}{r^2}\grave{R}(r) = 0. \tag{11.84}$$

Once this equation is solved for $\grave{R}(r)$, the function $R(\rho)$ is obtained using Equation 11.83. This result shows that the α parameter acts like a scaling factor for the solution, because the qualitative behavior of $R(\rho)$ does not depend on α. The ν parameter, however, is not just a scaling factor. The qualitative behavior of $\grave{R}(r)$, and thus $R(\rho)$, depends on the specific value of ν.

Bessel Functions of the First and Second Kind   Equation 11.84 can be solved using the method of Frobenius, described in the previous chapter. Recall that, with this method, we look for series solutions of the form

$$\grave{R}(r) = \sum_{n=0}^{\infty} c_n r^{n+s}. \tag{11.85}$$

By plugging this into Equation 11.84, and requiring $c_0 \neq 0$, we get the indicial equation for s:

$$s^2 - \nu^2 = 0 \tag{11.86}$$

with the result

$$s = \pm\nu. \tag{11.87}$$

This gives two solutions for $\grave{R}(r)$:

$$\grave{R}_1(r) = \sum_{n=0}^{\infty} a_n r^{n+\nu} \quad \text{from } s = +\nu \tag{11.88}$$

and

$$\grave{R}_2(r) = \sum_{n=0}^{\infty} b_n r^{n-\nu} \quad \text{from } s = -\nu. \tag{11.89}$$

Two different sets of coefficients, a_n and b_n, have been used to distinguish between the solutions.

When the first of these two series, Equation 11.88, is inserted back into the differential equation, the coefficients can be easily evaluated. The series is

$$\grave{R}_1(r) = a_o 2^\nu \nu! \sum_{m=0}^{\infty} \frac{(-1)^m}{m!(m+\nu)!} \left(\frac{r}{2}\right)^{2m+\nu}, \tag{11.90}$$

where a_o is an undetermined coefficient. If a_o is chosen to be $1/2^\nu \nu!$, the RHS of Equation 11.90 is defined as the νth-order Bessel function of the first kind:

$$J_\nu(r) \equiv \sum_{m=0}^{\infty} \frac{(-1)^m}{m!(m+\nu)!} \left(\frac{r}{2}\right)^{2m+\nu}. \tag{11.91}$$

Unfortunately, if we try the same trick with the series in Equation 11.89, we run into trouble. If ν is an integer, which is the most common situation, the b_n coefficients will always diverge after some finite number of terms. Consequently, when ν is an integer we cannot generate a second solution using this method. This divergence stems from a singularity of the original differential equation at $r = 0$. This second solution has a "logarithmic" singularity at $r = 0$, which means it cannot be represented by a power series. However, we can find the second solution if we modify the method of Frobenius slightly, by explicitly identifying the singularity with a logarithm factor. Now we look for a solution of the form,

$$\mathring{R}_2(r) = \ln(r)J_\nu(r) + \sum_{n=0}^{\infty} b_n r^{n+s}, \tag{11.92}$$

where $J_\nu(r)$ is the first solution we found in Equation 11.91. Actually, because the differential equation is linear, any linear combination of $\mathring{R}_1(r) \equiv J_\nu(r)$ and the $\mathring{R}_2(r)$ obtained with this method is an equally legitimate second solution. This gives us some flexibility in choosing the second solution. A common choice is called the Bessel function of the second kind, or sometimes the Neumann function of order ν. Deriving the result is straightforward for the $\nu = 0$ case. The general case for arbitrary ν, however, is quite complicated. We will simply give the result here and recommend you investigate the classic reference by Watson, *A Treatise on the Theory of Bessel Functions*, for the details. The second solution is

$$Y_\nu(r) = \frac{2}{\pi} J_\nu(r) \left(\ln \frac{r}{2} + \gamma\right)$$

$$+ \sum_{n=0}^{\infty} \frac{(-1)^{n+1}(h_n + h_{n+\nu})}{\pi n!(n+\nu)!} \left(\frac{r}{2}\right)^{2n+\nu}$$

$$- \sum_{n=0}^{\nu-1} \frac{(\nu-n-1)!}{\pi n!} \left(\frac{r}{2}\right)^{2n-\nu}, \tag{11.93}$$

where h_n is the partial harmonic series defined by

$$h_n = \begin{cases} 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} & n > 0 \\ 0 & n = 0 \end{cases}. \tag{11.94}$$

and γ is

$$\gamma \equiv \lim_{n \to \infty}(h_n - \ln n) \approx .5772, \tag{11.95}$$

which is called the Euler constant.
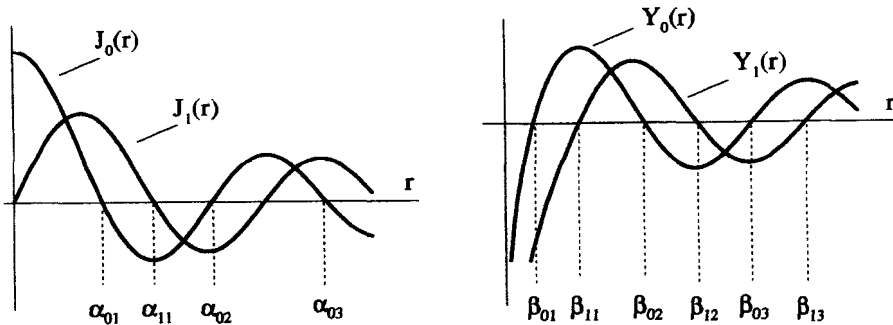
Figure 11.11   Bessel Functions of the First and Second Kind

Because these functions are so complicated, they have been extensively tabulated in reference books, such as the *Handbook of Mathematical Functions* by Abramowitz and Stegun. The $\nu = 0$ and $\nu = 1$ versions of these functions are plotted in Figure 11.11. All the $J_\nu(r)$ functions are zero at the origin, except for J_0, which has $J_0(0) = 1$. The functions all have the form of a damped oscillation, and the points where they cross zero turn out to be very important. These points are not evenly spaced, and we usually define the nth zero crossing of $J_\nu(r)$ as the special constant $\alpha_{\nu n}$. These values are also extensively tabulated in the literature. The $Y_\nu(r)$ functions all go to minus infinity as $r \to 0$. Away from the origin, these functions are again in the form of damped oscillations. The nth zero crossing of $Y_\nu(r)$ occurs at $r = \beta_{\nu n}$.

After all this, it must be remembered that we originally were seeking the general solution of $R(\rho)$ in Equation 11.81. We have obtained two independent solutions for $\dot{R}(r)$ in Equation 11.84. These solutions can be related by to $R(\rho)$ via Equation 11.83. Therefore, the general solution for $R(\rho)$ is any linear combination of the two solutions:

$$R(\rho) \to \begin{cases} J_\nu(\alpha\rho) \\ Y_\nu(\alpha\rho) \end{cases} . \tag{11.96}$$

The General Solution for $\Phi(\rho, \theta, z)$ The general solution for $\Phi(\rho, \theta, z)$, for the case where $c_z = \alpha^2 > 0$ and $c_\theta = -\nu^2 \leq 0$, is a linear sum of all the possible products of the separated functions,

$$\Phi(\rho, \theta, z) = \sum_\nu \sum_\alpha \begin{cases} \cos(\nu\theta) \\ \sin(\nu\theta) \end{cases} \begin{cases} \cosh(\alpha z) \\ \sinh(\alpha z) \end{cases} \begin{cases} J_\nu(\alpha\rho) \\ Y_\nu(\alpha\rho) \end{cases} , \tag{11.97}$$

where the summation is over all the allowed values of ν and α. In general, the boundary conditions will limit these possible values, as you will see in the example that follows.

Keep in mind, we could have just as easily chosen to write the θ-dependence using complex exponentials, but that is not as convenient when we expect $H(\theta)$ to be real. Likewise, the z-dependence has been written with the sinh and cosh

functions, but could just as easily have been expressed with growing or decaying exponential functions. Computational convenience dictates which choice makes more sense. Notice the θ-dependence is periodic, while the z-dependence is exponential. This confirms the general rule for separated solutions we declared at the beginning of the chapter. If a solution oscillates in one direction, it must exponentiate in another.

Applying the Boundary Conditions   Consider the boundary conditions described in Figure 11.12. The potential is controlled on the surface of a cylinder of radius r_o and length $2z_o$. The sides of the cylinder are held at $\Phi(r_o, \theta, z) = 0$. The top is held at $\Phi(\rho, \theta, z_o) = f_T(\rho, \theta)$, a potential that varies with both ρ and θ, while the bottom is held at a different potential, $\Phi(\rho, \theta, -z_o) = f_B(\rho, \theta)$. The problem is to determine the potential inside the cylinder.

It seems appropriate to start with the general solution given by Equation 11.97, because the range of θ is $0 < \theta < 2\pi$, which requires the solution to be periodic in θ. Also, given the potentials on the top and bottom, the solution does not appear to be periodic in z.

The geometry, boundary conditions, and a little physical reasoning allow the general form to be simplified somewhat. First, it would not make any physical sense for the potential to diverge as $\rho \to 0$, so the $R(\rho)$ solution will not have any of the $Y_\nu(\alpha\rho)$ solutions. Next, the boundary condition which forces $\Phi(r_o, \theta, z) = 0$ limits the possible values of α to a discrete set. Remember, the zeros of a particular $J_\nu(\alpha\rho)$ Bessel function occur when the argument is equal to one of the constants $\alpha_{\nu n}$. So to have $R(r_o) = 0$, we must have

$$\alpha r_o = \alpha_{\nu n}, \tag{11.98}$$

where $\nu = 0, 1, 2, \ldots$ and $n = 1, 2, 3, \ldots$. In this way, the summation over α in Equation 11.97 is converted to a summation over n, and the general solution is
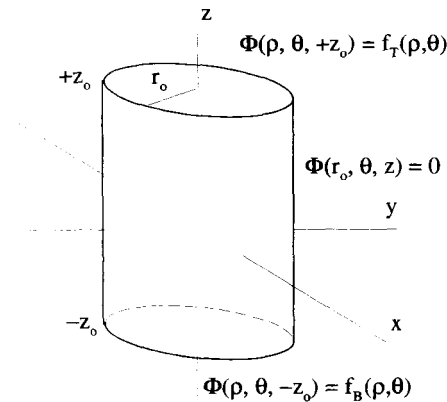
Figure 11.12   Cylindrical Boundary Conditions Nonperiodic in the z-Direction

reduced to

$$\Phi(\rho, \theta, z) = \sum_{\nu=0}^{\infty} \sum_{n=1}^{\infty} \begin{Bmatrix} \cos(\nu\theta) \\ \sin(\nu\theta) \end{Bmatrix} \begin{Bmatrix} \cosh(\alpha_{\nu n} z/r_o) \\ \sinh(\alpha_{\nu n} z/r_o) \end{Bmatrix} \begin{Bmatrix} J_\nu(\alpha_{\nu n}\rho/r_o). \end{Bmatrix} \qquad (11.99)$$

Notice that the value of α in the argument of the hyperbolic functions is determined by a boundary condition associated with the Bessel function.

If the shorthand notation on the RHS of Equation 11.99 is expanded, it is necessary to introduce four amplitude constants. In order to determine these constants, the top and bottom boundary conditions, yet unused, need to be introduced. These amplitude constants can be determined much like we found the amplitude constants in the Cartesian problems at the beginning of the chapter. By applying orthogonality conditions, we can isolate and evaluate one of the constants associated with a particular ν and a particular n. This will require two different orthogonality relations, one to isolate values of ν, and one to isolate values of n.

Orthogonality Relations We already know that the sin and cos functions which describe the θ-dependence obey an orthogonality condition, while the sinh and cosh functions do not. To evaluate all the amplitude constants, we need an additional orthogonality relation, which we can obtain from the Bessel functions. We can derive it by putting Bessel's equation in the form of an eigenvalue, eigenfunction problem, identify a Hermitian operator, and then determine the orthogonality properties of its eigenfunctions.

Originally, the $J_\nu(\alpha_{\nu n}\rho/r_o)$ Bessel functions were constructed to satisfy Equation 11.81:

$$\left[\frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial}{\partial \rho} - \frac{\nu^2}{\rho^2} \right] J_\nu(\alpha_{\nu n}\rho/r_o) = -\frac{\alpha_{\nu n}^2}{r_o^2} J_\nu(\alpha_{\nu n}\rho/r_o). \qquad (11.100)$$

This is in the form of a standard eigenvalue, eigenfunction problem with the weighting function $w(x) = 1$,

$$\mathcal{L}_{op}\phi_n = \lambda_n \phi_n, \qquad (11.101)$$

where

$$\mathcal{L}_{op} = \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial}{\partial \rho} - \frac{\nu^2}{\rho^2} \qquad (11.102)$$

$$\phi_n = J_\nu(\alpha_{\nu n}\rho/r_o), \qquad (11.103)$$

and

$$\lambda_n = -\frac{\alpha_{\nu n}^2}{r_o^2}. \qquad (11.104)$$

Because ν is inside the operator and fixed, different eigenfunctions of this operator are obtained by incrementing n, i.e., $n = 1, 2, 3, \ldots$.

These functions will obey an orthogonality condition if, when they are combined with $\mathcal{L}_{op}$, they satisfy the Hermitian condition, that is, if

$$\int_\Omega d\rho \ J_\nu(\alpha_{\nu m}\rho/r_o)\mathcal{L}_{op}J_\nu(\alpha_{\nu n}\rho/r_o)$$

equals

$$\int_\Omega d\rho \ J_\nu(\alpha_{\nu n}\rho/r_o)\mathcal{L}_{op}J_\nu(\alpha_{\nu m}\rho/r_o).$$

Unfortunately, it turns out that these two integrals are not equal and the Bessel functions, by themselves, are not orthogonal.

We know, from our proof in the previous section, that an operator in Sturm-Liouville form, combined with the appropriate boundary conditions, will automatically lead to the Hermitian condition. The operator in Equation 11.102 is not, itself, in Sturm-Liouville form, but if Equation 11.100 is multiplied by ρ it becomes

$$\left[\rho\frac{\partial^2}{\partial\rho^2} + \frac{\partial}{\partial\rho} - \frac{\nu^2}{\rho}\right]J_\nu(\alpha_{\nu m}\rho/r_o) = -\frac{\alpha_{\nu m}^2\rho}{r_o^2}J_\nu(\alpha_{\nu m}\rho/r_o), \qquad (11.105)$$

which can be written as

$$\left[\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho} - \frac{\nu^2}{\rho}\right]J_\nu(\alpha_{\nu m}\rho/r_o) = -\frac{\alpha_{\nu m}^2\rho}{r_o^2}J_\nu(\alpha_{\nu m}\rho/r_o). \qquad (11.106)$$

The modified operator for this equation is

$$\mathcal{L}_{op}' = \frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho} - \frac{\nu^2}{\rho}, \qquad (11.107)$$

which is in Sturm-Liouville form. Equation 11.106 is in the form of a eigenfunction equation, with a weighting function of $w(\rho) = \rho$:

$$\mathcal{L}_{op}'\phi_n = \lambda_n\rho\phi_n. \qquad (11.108)$$

Here again, the eigenfunctions and eigenvalues are

$$\phi_n = J_\nu(\alpha_{\nu n}\rho/r_o) \qquad (11.109)$$

$$\lambda_n = -\frac{\alpha_{\nu n}^2}{r_o^2}. \qquad (11.110)$$

Because the operator $\mathcal{L}_{op}'$ is in Sturm-Liouville form, the Hermitian condition will be satisfied if Equation 11.57 is satisfied:

$$\left[J_\nu(\alpha_{\nu m}\rho/r_o)\rho\frac{dJ_\nu(\alpha_{\nu n}\rho/r_o)}{d\rho}\right]\bigg|_\Omega = 0. \qquad (11.111)$$

Because the range of interest for this problem has an obvious lower limit of $\rho = 0$ and an upper limit of $\rho = r_o$, this condition is easily satisfied. The lower limit vanishes due to the presence of the ρ term, and the upper limit vanishes because $J_\nu(\alpha_{\nu n})$ is by definition zero. Thus with Ω covering the range $0 < \rho < r_o$, the modified operator of Equation 11.107 is indeed Hermitian. This means the Bessel functions form a complete set of functions and obey the orthogonality condition of Equation 11.49:

$$\int_0^{r_o} d\rho \, J_\nu(\alpha_{\nu n}\rho/r_o)\rho J_\nu(\alpha_{\nu m}\rho/r_o) = 0 \qquad n \neq m. \tag{11.112}$$

In order to perform expansions, we need one more tool. We need to also know the value of the integral of Equation 11.112 when $m = n$. We give the result, without the proof, that the general expression is

$$\int_o^{r_o} d\rho \, J_\nu(\alpha_{\nu n}\rho/r_o)\rho J_\nu(\alpha_{\nu m}\rho/r_o) = \delta_{nm}\frac{r_o^2}{2}J_{\nu+1}^2(\alpha_{\nu n}), \tag{11.113}$$

where we have used the Kronecker delta symbol, δ_{nm}.

The orthogonality condition of Equation 11.113 now allows any function $f(\rho)$, which obeys $f(\rho = r_o) = 0$ and $f(\rho) < \infty$, to be expanded in the range $0 \leq \rho \leq r_o$ using the Bessel functions. In other words, we can write

$$f(\rho) = \sum_{n=1}^{\infty} A_n J_\nu(\alpha_{\nu n}\rho/r_o), \tag{11.114}$$

with

$$A_n = \frac{\int_o^{r_o} d\rho \, f(\rho)\rho J_\nu(\alpha_{\nu n}\rho/r_o)}{(r_o^2/2)J_{\nu+1}^2(\alpha_{\nu n})}. \tag{11.115}$$

Evaluating the Amplitude Constants   Now we can pick up the cylindrical problem at the point of Equation 11.99. The problem is further simplified if we impose odd symmetry in z by setting $f_T(\rho, \theta) = -f_B(\rho, \theta)$. This eliminates the hyperbolic cosine functions and Equation 11.99 becomes

$$\Phi(\rho, \theta, z) = \sum_{\nu=0}^{\infty}\sum_{n=1}^{\infty} \begin{Bmatrix} \cos(\nu\theta) \\ \sin(\nu\theta) \end{Bmatrix} \begin{Bmatrix} \sinh(\alpha_{\nu n}z/r_o)J_\nu(\alpha_{\nu n}\rho/r_o). \end{Bmatrix} \tag{11.116}$$

At this point, it is convenient to insert the amplitude constants and explicitly write out the summations of Equation 11.116:

$$\Phi(\rho, \theta, z) = \sum_{\nu=0}^{\infty}\sum_{n=1}^{\infty} \begin{Bmatrix} A_{\nu n}\sin(\nu\theta)\sinh(\alpha_{\nu n}z/r_o)J_\nu(\alpha_{\nu n}\rho/r_o) \\ +B_{\nu n}\cos(\nu\theta)\sinh(\alpha_{\nu n}z/r_o)J_\nu(\alpha_{\nu n}\rho/r_o) \end{Bmatrix}. \tag{11.117}$$

Now we need to determine the $A_{\nu n}$ and $B_{\nu n}$ coefficients using the boundary conditions on the top and bottom surfaces. Because the odd symmetry in z has already

been imposed, we can use either the top or bottom boundary. Picking the top, the series expression is evaluated at $z = +z_o$:

$$f_T(\rho, \theta) = \sum_{\nu=0}^{\infty} \sum_{n=1}^{\infty} \left\{ \begin{array}{l} A_{\nu n} \sin(\nu\theta) \sinh(\alpha_{\nu n} z_o/r_o) J_\nu(\alpha_{\nu n}\rho/r_o) \\ +B_{\nu n} \cos(\nu\theta) \sinh(\alpha_{\nu n} z_o/r_o) J_\nu(\alpha_{\nu n}\rho/r_o) \end{array} \right\}. \qquad (11.118)$$

The amplitude constants can now be isolated and evaluated by using the orthogonality conditions of the trigonometric functions and of the Bessel functions. Because there are two summations, one over n and one over ν, two orthogonality conditions are required.

To evaluate the $A_{\nu n}$, operate on both sides of Equation 11.118 with

$$\int_o^{2\pi} d\theta \, \sin(\nu_o \theta). \qquad (11.119)$$

This filters out all but the $\nu = \nu_o$ sine terms in Equation 11.118:

$$\int_o^{2\pi} d\theta \, \sin(\nu_o \theta) f_T(\rho, \theta) = \sum_{n=1}^{\infty} A_{\nu_o n} \pi \sinh(\alpha_{\nu_o n} z/r_o) J_{\nu_o}(\alpha_{\nu_o n}\rho/r_o). \qquad (11.120)$$

Notice how the order of the Bessel function ν_o has been selected, not by the orthogonality of the Bessel functions, but by the orthogonality of the sine functions.

A particular value of n is selected by operating on both sides of Equation 11.120 with

$$\int_0^{r_o} d\rho \, \rho J_{\nu_o}(\alpha_{\nu_o n_o}\rho/r_o). \qquad (11.121)$$

This singles out the n_o term to give

$$\int_0^{r_o} d\rho \, \rho J_{\nu_o}(\alpha_{\nu_o n_o}\rho/r_o) \int_o^{2\pi} d\theta \, \sin(\nu_o\theta) f_T(\rho, \theta)$$

$$= A_{\nu_o n_o} \pi \sinh(\alpha_{\nu_o n_o} z_o/r_o) \frac{r_o^2}{2} J_{\nu_o+1}^2(\alpha_{\nu_o n_o}) \qquad (11.122)$$

so that, in general,

$$A_{\nu n} = \frac{\int_0^{r_o} d\rho \, \rho J_\nu(\alpha_{\nu n}\rho/r_o) \int_o^{2\pi} d\theta \, \sin(\nu\theta) f_T(\rho, \theta)}{(\pi r_o^2/2) \sinh(\alpha_{\nu n} z_o/r_o) J_{\nu+1}^2(\alpha_{\nu n})}. \qquad (11.123)$$

A similar operation determines the $B_{\nu n}$

$$B_{\nu n} = \frac{\int_0^{r_o} d\rho \, \rho J_\nu(\alpha_{\nu n}\rho/r_o) \int_o^{2\pi} d\theta \, \cos(\nu\theta) f_T(\rho, \theta)}{(\pi r_o^2/2) \sinh(\alpha_{\nu n} z_o/r_o) J_{\nu+1}(\alpha_{\nu n})} \quad \nu > 0 \qquad (11.124)$$

and

$$B_{0n} = \frac{\int_o^{r_o} d\rho\, \rho J_0(\alpha_{0n}\rho/r_o) \int_0^{2\pi} d\theta\, f_T(\rho, \theta)}{(\pi r_o^2) \sinh(\alpha_{0n}z_o/r_o)J_1^2(\alpha_{0n})} \qquad \nu = 0. \qquad (11.125)$$

These amplitude constants are then inserted into Equation 11.117 to complete the solution.

A simpler result is obtained if it is assumed that the potential on the top and bottom is not a function of θ, i.e., $f_T(\rho, \theta) \to f_t(\rho)$. In a case like this, only $\nu = 0$ is allowed. There are no A_{0n} terms because $\sin(0) = 0$. Only B_{0n} amplitudes are present, and the solution for $\Phi(\rho, z)$ inside the cylinder becomes

$$\Phi(\rho, z) = \sum_{n=1}^{\infty} B_{0n} \sinh(\alpha_{0n}z/r_o)J_0(\alpha_{0n}\rho/r_o), \qquad (11.126)$$

with

$$B_{0n} = \frac{\int_0^{r_o} d\rho\, \rho J_o(\alpha_{0n}\rho/r_o)f_T(\rho)}{(r_o^2/2) \sinh(\alpha_{0n}z_o/r_o)J_1^2(\alpha_{0n})}. \qquad (11.127)$$

11.3.3 Solutions for $R(\rho)$ with $c_z < 0$ and $c_\theta \leq 0$

We now turn our attention to cylindrical solutions of Laplace's equation that oscillate in both the z- and θ-directions. The general solution is still a product of the separated solutions

$$\Phi(\rho, \theta, z) = R(\rho)H(\theta)Z(z), \qquad (11.128)$$

and the functions $Z(z)$ and $H(\theta)$ still must satisfy Equations 11.70 and 11.76, respectively. Now, however, $c_z = -k^2 < 0$ and $c_\theta = -\nu^2 \leq 0$. This means that the form of solutions for these equations will be

$$Z(z) \to \begin{cases} e^{ikz} \\ e^{-ikz} \end{cases} \quad \text{or} \quad Z(z) \to \begin{cases} \sin(kz) \\ \cos(kz) \end{cases} \qquad (11.129)$$

and

$$H(\theta) \to \begin{cases} e^{i\nu\theta} \\ e^{-i\nu\theta} \end{cases} \quad \text{or} \quad H(\theta) \to \begin{cases} \sin(\nu\theta) \\ \cos(\nu\theta) \end{cases}. \qquad (11.130)$$

The choice between the trigonometric and exponential forms of the solutions is, as usual, determined by whichever is more convenient for a particular set of boundary conditions.

The separated equation for $R(\rho)$ becomes

$$\left[\frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} - k^2 - \frac{\nu^2}{\rho^2} \right] R(\rho) = 0. \qquad (11.131)$$

Equation 11.131 is identical to Equation 11.81 with $\alpha^2 \rightarrow -k^2$ or $\alpha \rightarrow ik$. The solutions to Equation 11.81 were found to be Bessel functions of the first and second kind:

$$J_\nu(\alpha\rho) \quad \text{and} \quad Y_\nu(\alpha\rho).$$

The solutions to Equation 11.131 can therefore be written as

$$R(\rho) \rightarrow \begin{cases} J_\nu(ik\rho) \\ Y_\nu(ik\rho) \end{cases}. \tag{11.132}$$

The imaginary arguments in the Bessel functions have a similar effect as imaginary arguments in the sine and cosine functions. These functions no longer oscillate, but they grow or decay with ρ. We might have expected this, because the general rule for the separated solutions of Laplace's equation in Cartesian systems was that at least one separated solution was growing or decaying.

The functions in Equation 11.132 are not in the most convenient form, however, because they are not pure real functions of ρ. We can correct this problem by constructing two different solutions which are pure real:

$$I_\nu(k\rho) = i^{-\nu} J_\nu(ik\rho) \tag{11.133}$$

$$K_\nu(k\rho) = (\pi/2)i^{\nu+1}[J_\nu(ik\rho) + iY_\nu(ik\rho)]. \tag{11.134}$$

These two functions are called modified Bessel functions of the first and second kind. Again, like the Bessel functions, the modified Bessel functions are well tabulated in the literature. Graphs of the first few, $I_0(r)$, $I_1(r)$, $K_0(r)$ and $K_1(r)$, are shown in Figure 11.13. Notice, the $I_\nu(r)$ functions are finite at $r = 0$ and blow up as $r \rightarrow \infty$. In contrast, the $K_\nu(r)$ functions blow up as $r \rightarrow 0$ and go to zero as $r \rightarrow \infty$.

Now, we can write the general solution of $R(\rho)$ as

$$R(\rho) \rightarrow \begin{cases} I_\nu(k\rho) \\ K_\nu(k\rho) \end{cases}, \tag{11.135}$$

and the general form of $\Phi(\rho, \theta, z)$ for this case becomes

$$\Phi(\rho, \theta, z) = \sum_\nu \sum_K \begin{cases} \cos(\nu\theta) \\ \sin(\nu\theta) \end{cases} \begin{cases} \cos(kz) \\ \sin(kz) \end{cases} \begin{cases} I_\nu(k\rho) \\ K_\nu(k\rho) \end{cases} \tag{11.136}$$
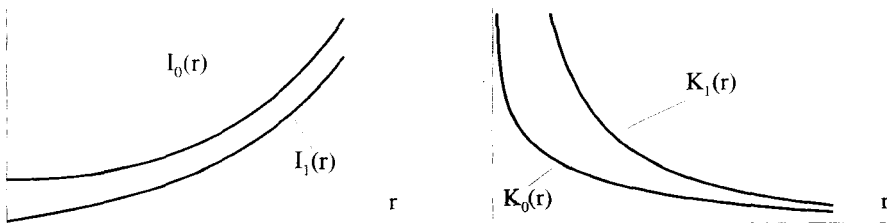
Figure 11.13 Modified Bessel Functions of the First and Second Kind

In this expression, sine and cosine functions have been used for the z- and θ- dependencies. As always, complex exponential functions could have been used instead, but the sine-cosine functions are better when we expect the final answer to be pure real. Again notice the presence of at least one oscillating factor and one growing/decaying factor in each term.

Example 11.4   As an example of an electrostatic problem that uses the type of solution given in Equation 11.136, consider the geometry shown in Figure 11.14. This is quite similar to the problem discussed in Figure 11.7, which described a periodic boundary condition in Cartesian coordinates. In this problem, the potential is controlled on the $\rho = r_o$ surface of the cylinder. Its value is $+V_o$ for $0 < z < L/2$ and $-V_o$ for $L/2 < z < L$. We are interested in a solution for $\Phi(r_o, \theta, z)$ inside the cylinder where $\rho < r_o$.

Notice that the boundary condition is periodic, so $\Phi(r_o, \theta, z) = \Phi(r_o, \theta, z + L)$ for all z. Consequently, the solution for this problem must also be periodic in z, and so the solution to Laplace's equation for $\rho < r_o$ must fall into the general form of Equation 11.136. The symmetry of these particular boundary conditions allows us to simplify this general form tremendously. First of all, there is no θ-variation, so we can set $H(\theta) = 1$, and the sum over ν only uses the $\nu = 0$ term. Second, the solution must be finite for $\rho = 0$, so there are no $K_0(k\rho)$ terms, because they diverge as $\rho \to 0$. Next, because the boundary conditions have odd symmetry about z, all the $\cos(kz)$ terms vanish, because they are all even functions. Finally, the periodicity of the boundary conditions forces k to be limited to the discrete set of values:

$$k \to k_n = \frac{2\pi n}{L} \quad \text{where } n = 0, 1, 2, \ldots. \tag{11.137}$$

With these simplifications, Equation 11.136 reduces to

$$\Phi(\rho, \theta, z) = \sum_{n=1}^{\infty} A_n \sin(k_n z) I_0(k_n \rho) \tag{11.138}$$

$$= \sum_{n=1}^{\infty} A_n \sin\left(\frac{2\pi n z}{L}\right) I_0\left(\frac{2\pi n \rho}{L}\right). \tag{11.139}$$

Notice the value of k appears not only as an argument of the $\sin(kz)$ term, but also in the argument of the modified Bessel function.

As always, the A_n coefficients can be determined using an orthogonality condition and the boundary conditions at $\rho = r_o$. Operate on both sides of Equation 11.138 by

$$\int_0^L dz \, \sin\left(\frac{2\pi m z}{L}\right)$$

to obtain

$$\int_0^L dz \, \sin\left(\frac{2\pi m z}{L}\right) \Phi(r_o, z) = \frac{A_m L}{2} I_0\left(\frac{2\pi m r_o}{L}\right). \tag{11.140}$$
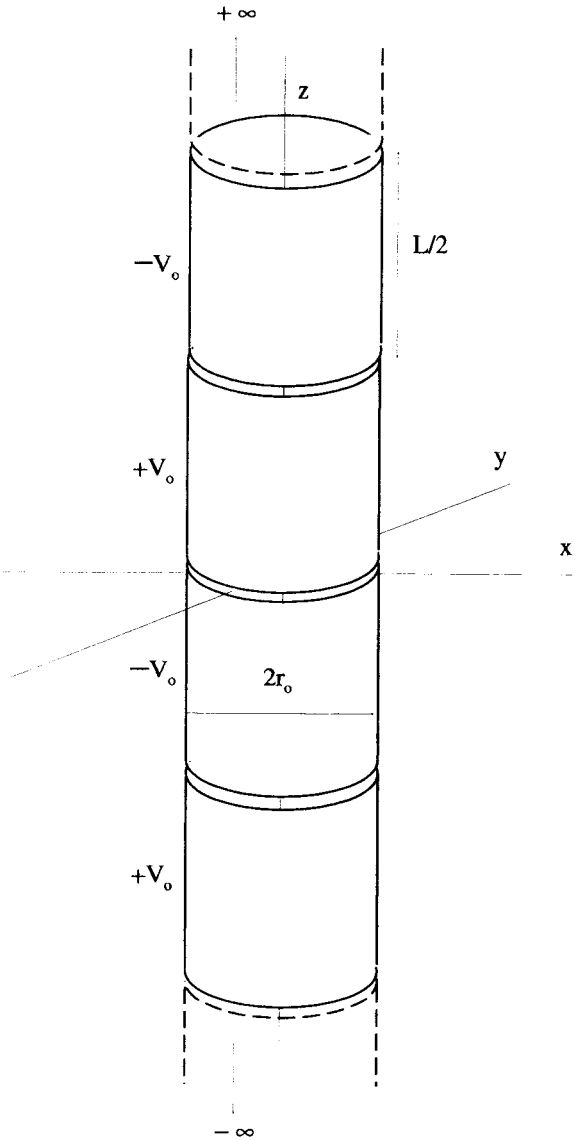
Figure 11.14   Cylindrical Laplace Problem Periodic in the z-Direction

Now, if we require

$$\Phi(r_o, z) = \begin{cases} +V_o & 0 < z < L/2 \\ -V_o & L/2 < z < L \end{cases} \tag{11.141}$$

and perform the integral on the LHS of Equation 11.140, the A_n amplitudes evaluate to

$$A_n = \begin{cases} 0 & n \text{ even} \\ \dfrac{4V_o}{n\pi I_0(2\pi n r_o/L)} & n \text{ odd} \end{cases} . \qquad (11.142)$$

The final solution to this problem for $\rho < r_o$ becomes

$$\Phi(\rho, \theta, z) = \sum_{n=1,3,5,\dots}^{\infty} \frac{4V_o}{n\pi I_0(2\pi n r_o/L)} \sin\left(\frac{2\pi n z}{L}\right) I_0\left(\frac{2\pi n \rho}{L}\right) . \qquad (11.143)$$

In this example, the modified Bessel functions did not combine with the ρ-dependent operator to satisfy a Hermitian condition. Consequently, they were not orthogonal functions and could not be used to select a particular value of n. Instead, we used the orthogonality of the sine functions to perform the expansion.

Example 11.5 As a final example, consider the cylindrical boundary value problem shown in Figure 11.15. The potential is held at zero on the top and bottom and at $+V_o$ on the sides. The solution to Laplace's equation is sought for $\rho < r_o$ and $0 < z < L/2$.

The first thing to decide is the form of the general solution. Is it going to be Equation 11.97, with oscillating solutions in ρ and nonoscillating solutions in z, or Equation 11.136, with oscillating solutions in z and nonoscillating solutions in ρ? The nonoscillating solutions in z clearly cannot work, because if the amplitudes are adjusted to make $\Phi = 0$ at $z = 0$, this solution cannot also be zero at $z = L/2$. Therefore, we must choose the form of Equation 11.136.

Now that we have determined that the solutions must be periodic in z, what is the spatial period? Remember we tackled a similar problem in Cartesian geometry, where we successfully converted a nonperiodic problem into a periodic one. Let's try the same process here. We can construct an infinite cylinder, of which the region shown in Figure 11.15 is but a single "cell." The zero potentials at the top and bottom
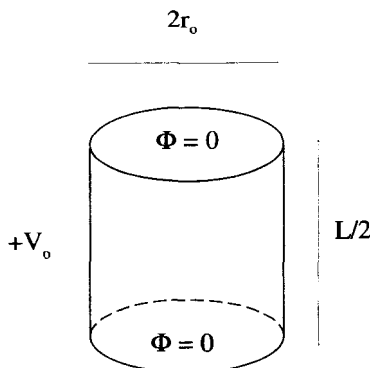
Figure 11.15 Boundary Conditions for a Cylindrical Can
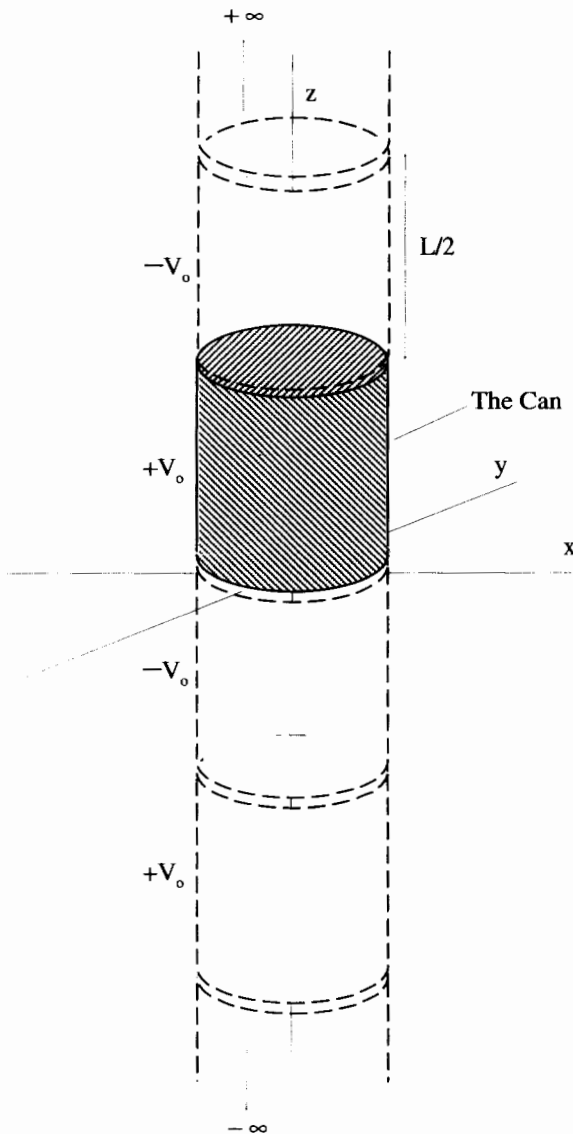
Figure 11.16 The Cylindrical Can Converted to a Periodic Problem

of this cell must be established by the periodic potential applied to the $\rho = r_o$ surface above and below this cell. This situation is shown in Figure 11.16.

From Figure 11.16 it is clear that the solution inside the can will be identical to Equation 11.143, the solution obtained in the previous example. Of course, in this case the solution is only valid for $0 < z < L/2$.

11.4 SPHERICAL SOLUTIONS

The solutions to Laplace's equation in spherical geometry are fairly complicated, but they are important to understand because of their frequent use in quantum mechanics. The organization of the periodic table of elements is due primarily to the nature of these types of solutions. We will take the same approach we did with the other coordinate systems and look for separable solutions.

Recall, the spherical coordinates (r, θ, ϕ) of a point P are defined as shown in Figure 11.17. The coordinates are limited to the ranges $0 < r < +\infty$, $0 < \theta < \pi$, and $0 < \phi < 2\pi$. The Laplace operator in this system becomes

$$\nabla^2 = \frac{1}{r}\frac{\partial^2}{\partial r^2}r + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2}{\partial\phi^2}. \tag{11.144}$$

Again, we seek solution to

$$\nabla^2\Phi(r, \theta, \phi) = 0 \tag{11.145}$$

by the method of separation of variables. However, because of the r-factor in the first term of Equation 11.144, it turns out to be more convenient to include a $1/r$ factor in the separation process. That is, we will look for a solution with the form:

$$\Phi(r, \theta, \phi) = \frac{R(r)}{r}P(\theta)F(\phi). \tag{11.146}$$

Substituting this into Laplace's equation and multiplying by

$$\frac{r^3 \sin^2\theta}{R(r)P(\theta)F(\phi)},$$

we obtain

$$\frac{r^2 \sin^2\theta}{R(r)}\frac{\partial^2 R(r)}{\partial r^2} + \frac{\sin\theta}{P(\theta)}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial P(\theta)}{\partial\theta} + \frac{1}{F(\phi)}\frac{\partial^2 F(\phi)}{\partial\phi^2} = 0. \tag{11.147}$$
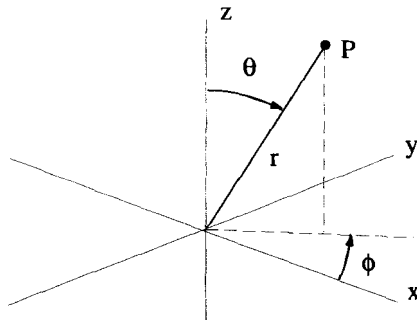
Figure 11.17 The Spherical Coordinates

Separating the ϕ-Dependence   Because the last term on the LHS of Equation 11.147 contains all the ϕ-dependence, it can be set equal to a separation constant. In the most common spherical problems, the range of ϕ is $0 < \phi < 2\pi$, and to keep the potential single-valued, we must have $\Phi(r, \theta, \phi) = \Phi(r, \theta, \phi + 2\pi)$. This means the $F(\phi)$ solutions must be periodic, and thus the separation constant cannot be positive. We will stick with the convention used by most scientists and define this separation constant as

$$\frac{1}{F(\phi)} \frac{\partial^2 F(\phi)}{\partial \phi^2} = -m^2 \leq 0. \tag{11.148}$$

We have discussed the solutions to this equation already. $F(\phi)$ can take either of two forms:

$$F(\phi) \rightarrow \begin{cases} e^{+im\phi} \\ e^{-im\phi} \end{cases} \quad \text{or} \quad F(\phi) \rightarrow \begin{cases} \sin(m\phi) \\ \cos(m\phi) \end{cases}. \tag{11.149}$$

In electrostatics, where the solutions need to be pure real, the second form of Equation 11.149 is usually preferred. In quantum mechanics, the solutions can be complex, and the first form is frequently used. Regardless of the form, the periodicity over the range $0 < \phi < 2\pi$ forces m to take on the discrete values

$$m = 0, 1, 2, \ldots. \tag{11.150}$$

Separating the r-Dependence   Now we can proceed to separate the r-dependence. Substituting $-m^2$ for the ϕ-dependent term, Equation 11.147 reduces to

$$\frac{r^2 \sin^2 \theta}{R(r)} \frac{\partial^2 R(r)}{\partial r^2} + \frac{\sin \theta}{P(\theta)} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial P(\theta)}{\partial \theta} - m^2 = 0. \tag{11.151}$$

After dividing this entire equation by $\sin^2 \theta$, the r-dependence is confined to the first term, which can be set equal to a second separation constant. For reasons that will become clear later, we will give this constant a rather peculiar form:

$$\frac{r^2}{R(r)} \frac{\partial^2 R(r)}{\partial r^2} = \ell(\ell + 1). \tag{11.152}$$

What is the general solution for $R(r)$? One could go through the entire method of Frobenius to generate series solutions for this equation, but in this case, it is not difficult to see that the series will collapse to a single term. If we look for a solution with the form

$$R(r) \propto r^n, \tag{11.153}$$

Equation 11.152 is satisfied if n satisfies the algebraic expression

$$(n)(n - 1) = \ell(\ell + 1). \tag{11.154}$$

This implies two solutions for n,

$$n = \ell + 1 \tag{11.155}$$

and

$$n = -\ell. \tag{11.156}$$

The solutions for $R(r)$ then are

$$R(r) \rightarrow \begin{cases} r^{-\ell} \\ r^{\ell+1} \end{cases}. \tag{11.157}$$

This can easily be confirmed by substituting these solutions back into Equation 11.152. At this point, we cannot say too much about these solutions, because we do not know if ℓ is real, imaginary, or complex. To determine the nature of ℓ, the θ-dependent equation must first be solved.

Separating the θ-Dependence Substituting $\ell(\ell + 1)$ for the r-dependent term in Equation 11.151 gives

$$\frac{1}{P(\theta)\sin\theta} \frac{\partial}{\partial\theta} \sin\theta \frac{\partial P(\theta)}{\partial\theta} - \frac{m^2}{\sin^2\theta} + \ell(\ell + 1) = 0. \tag{11.158}$$

This equation is called the associated Legendre equation and occurs often in quantum mechanics.

The trigonometric functions in Equation 11.158 are difficult to manipulate, but we can simplify things quite a bit by introducing the variable change

$$x = \cos\theta. \tag{11.159}$$

Because $0 \leq \theta \leq \pi$, the range for x is $-1 \leq x \leq +1$. The extreme values must be included in the respective ranges to include the z-axis. This turns out to be a very important fact when we begin to solve this equation. With this change of variables, $dx = -\sin\theta \, d\theta$ and $\sin^2\theta = 1 - x^2$, and Equation 11.158 becomes

$$\left\{ \frac{\partial}{\partial x} \left[(1 - x^2)\frac{\partial}{\partial x} \right] - \frac{m^2}{1 - x^2} + \ell(\ell + 1) \right\} \grave{P}(x) = 0. \tag{11.160}$$

This new function $\grave{P}(x)$ is related to our original function $P(\theta)$ by $P(\theta) = \grave{P}(\cos\theta)$. Notice Equation 11.160 is in Sturm-Liouville form. We can maneuver it to be in the form of an eigenvalue/eigenfunction equation as well:

$$\left\{ \frac{\partial}{\partial x} \left[(1 - x^2)\frac{\partial}{\partial x} \right] - \frac{m^2}{1 - x^2} \right\} \grave{P}(x) = -\ell(\ell + 1)\grave{P}(x). \tag{11.161}$$

Based on our previous experiences with equations in this form, we know the solutions will form a complete, orthogonal set.

Because of its quantum mechanical significance, we will present a fairly detailed solution to Equation 11.161. We will begin by looking for solutions that have no ϕ-dependence, i.e., solutions with $m = 0$ that are symmetric around the z-axis. Then we will tackle the general case.

11.4.1 Axially Symmetric Solutions with m = 0

With $m = 0$, the ϕ-dependence becomes

$$F(\phi) \to 1, \tag{11.162}$$

the r-dependence remains the same

$$\frac{R(r)}{r} \to \begin{cases} r^\ell \\ r^{-\ell-1} \end{cases}, \tag{11.163}$$

and the equation for $\grave{P}(x)$ becomes

$$\frac{\partial}{\partial x}\left[(1 - x^2)\frac{\partial \grave{P}(x)}{\partial x}\right] = -\ell(\ell + 1)\grave{P}(x). \tag{11.164}$$

At this point, it is clear that the solution for $\grave{P}(x)$ depends upon the choice of ℓ, so let's start labeling the functions with an ℓ index: $\grave{P}(x) \to \grave{P}_\ell(x)$. The solution to Equation 11.164 can be obtained using the method of Frobenius by setting

$$\grave{P}_\ell(x) = \sum_{n=0}^{\infty} a_n x^{n+s}. \tag{11.165}$$

With this substitution,

$$\frac{\partial \grave{P}_\ell(x)}{\partial x} = \sum_{n=0}^{\infty} a_n(n + s)x^{n+s-1} \tag{11.166}$$

and Equation 11.164 becomes

$$\sum_{n=0}^{\infty}\left\{\frac{\partial}{\partial x}\left[a_n(n + s)x^{n+s-1} - a_n(n + s)x^{n+s+1}\right] + \ell(\ell + 1)a_n x^{n+s}\right\} = 0. \tag{11.167}$$

Performing the second differentiation gives

$$\sum_{n=0}^{\infty}\left[a_n(n + s)(n + s - 1)x^{n+s-2} - a_n(n + s)(n + s + 1)x^{n+s}\right.$$

$$\left. + \ell(\ell + 1)a_n x^{n+s}\right] = 0. \tag{11.168}$$

The next step is to identify the coefficients of every power of x, and set them equal to zero:

$$a_0(s)(s - 1)x^{s-2} + a_1(1 + s)(s)x^{s-1} + \sum_{n=2}^{\infty}\{a_n(n + s)(n + s - 1) \tag{11.169}$$

$$- a_{n-2}[(n + s - 2)(n + s - 1) - \ell(\ell + 1)]\}x^{n+s-2} = 0.$$

The coefficients for the x^{s-2} and the x^{s-1} terms are individually set equal to zero to form the indicial equations: for s

$$a_0(s)(s-1) = 0 \qquad\qquad (11.170)$$

$$a_1(s+1)(s) = 0. \qquad\qquad (11.171)$$

Taking $a_0 \neq 0$, Equation 11.170 gives $s = 0$ and $s = 1$. These two values of s generate two solutions for $\grave{P}_\ell(x)$:

$$\grave{A}_\ell(x) = \sum_{n=0}^{\infty} a_n x^{n+1} \qquad \text{from } s = 1 \qquad\qquad (11.172)$$

and

$$\grave{B}_\ell(x) = \sum_{n=0}^{\infty} b_n x^n \qquad \text{from } s = 0. \qquad\qquad (11.173)$$

Two different symbols, a_n and b_n have been used for the coefficients to distinguish between the two different solutions.

The solution $\grave{P}_\ell(x) = \grave{A}_\ell(x)$ is generated with $s = 1$, $a_0 \neq 0$, and $a_1 = 0$. The solution $\grave{P}_\ell(x) = \grave{B}_\ell(x)$ is generated with $s = 0$, $b_0 \neq 0$, and $b_1 = 0$. It is worth mentioning that there are other possible choices. When $s = 1$, Equation 11.171 requires $a_1 = 0$, but with $s = 0$, Equation 11.171 can be satisfied with a nonzero b_1. So an alternate form of $\grave{B}_\ell(x)$ could be generated with both b_0 and b_1 nonzero. This new solution, however, would be a just a linear combination of our original $\grave{B}_\ell(x)$ and $\grave{A}_\ell(x)$. Also, the indicial equations could be satisfied with $a_0 = b_0 = 0$, and both a_1 and b_1 nonzero. In this case, Equation 11.171 would require different values for s, namely $s = -1$ and $s = 0$, but we would generate exactly the same solutions as our original choice. This makes sense, because the form of the series expansion in Equation 11.165 implicitly assumes that a_0 is the lowest nonzero term.

The last term in Equation 11.169 provides generating equations for a_n and b_n for $n \geq 2$:

$$a_n = a_{n-2} \left. \frac{(n+s-2)(n+s-1) - \ell(\ell+1)}{(n+s)(n+s-1)} \right|_{s=1} \qquad\qquad (11.174)$$

$$b_n = b_{n-2} \left. \frac{(n+s-2)(n+s-1) - \ell(\ell+1)}{(n+s)(n+s-1)} \right|_{s=0}. \qquad\qquad (11.175)$$

We will start off both series with $a_0 = 1$ and $b_0 = 1$.

These series have very interesting convergence properties. For $n \gg 1$, $a_n = a_{n-2}$ and $b_n = b_{n-2}$, so the ends of both series look like

$$\grave{A}_\ell(x) = \cdots + a_{n \gg 1} \left\{ x^n + x^{n+2} + x^{n+4} + \cdots \right\} \qquad\qquad (11.176)$$

and

$$\grave{B}_\ell(x) = \cdots + b_{n \gg 1} \left\{ x^n + x^{n+2} + x^{n+4} + \cdots \right\}. \qquad\qquad (11.177)$$

Both these solutions must be valid for $0 \leq \theta \leq \pi$, or equivalently $1 \geq x \geq -1$. From the ratio test on Equations 11.176 and 11.177, it is clear that the only way for these series to converge at the end points is to have them terminate, i.e., both a_n and b_n must be zero for some, not necessarily the same, value of n. So we must look for the conditions that allow these generating functions to start with nonzero a_0 (or b_0), but which eventually cause the series terminate.

Starting with the a_n, Equation 11.174 simplifies to

$$a_n = a_{n-2} \frac{(n-1)(n) - \ell(\ell+1)}{(n+1)(n)}. \tag{11.178}$$

Because $a_1 = 0$, this equation is evaluated for only even values of n, i.e., $n = 2, 4, 6, \ldots$. Now, if a_n is to be zero for some even value of n, the numerator must go to zero. That is to say, for some even value of n,

$$(n-1)(n) - \ell(\ell+1) = 0. \tag{11.179}$$

This will cause a_n and all subsequent $a_{m>n}$ to be zero. Similar arguments conclude that the series representation for $B_\ell(x)$ terminates at an even value of n given by

$$(n-2)(n-1) - \ell(\ell+1) = 0. \tag{11.180}$$

Because n is a positive, even integer, the only way either Equation 11.179 or 11.180 can be satisfied is if ℓ is a real, but not necessarily positive, integer. This is our first condition on the separation constant $\ell(\ell+1)$ that isolated the r-dependence. Now we need to determine what these terminating solutions for $\grave{P}_\ell(x)$ look like. We will individually determine different solutions as ℓ is indexed, i.e., $\ell = 0, \pm 1, \pm 2, \ldots$.

Starting with $\ell = 0$, the $A_0(x)$ series terminates when Equation 11.179 is satisfied, i.e., when

$$(n-1)(n) = 0 \quad \text{for } n = 2, 4, 6, \ldots. \tag{11.181}$$

It is clear that no positive, even value of n will satisfy this condition. Thus, when $\ell = 0$ the $\grave{A}_0(x)$ series of Equation 11.172 does not converge for $x = \pm 1$, and is therefore unacceptable if the solution must be valid on the z-axis (where $x = \pm 1$). With $\ell = 0$ the $\grave{B}_0(x)$ series terminates when Equation 11.180 is satisfied, i.e., when

$$(n-2)(n-1) = 0 \quad \text{for } n = 2, 4, 6, \ldots. \tag{11.182}$$

This equation is satisfied with $n = 2$. So with $\ell = 0$, the $\grave{B}_0(x)$ series of Equation 11.173 terminates and therefore converges for $x = \pm 1$ with $b_o = 1, b_1 = 0$, and $b_2 = 0$. Since the $\grave{A}_0(x)$ solution was unacceptable, the only terminating solution for $\ell = 0$ is

$$\grave{P}_0(x) = \grave{B}_0(x) = 1. \tag{11.183}$$

We will now work out the solution for $\ell = 1$ and a pattern for the solutions will become clear. Referring to Equation 11.179, the $\grave{A}_1(x)$ series terminates when

$$(n-1)(n) - 2 = 0 \quad \text{for } n = 2, 4, 6, \ldots. \tag{11.184}$$

The $\grave{B}_1(x)$ series terminates when

$$(n - 2)(n - 1) - 2 = 0 \quad \text{for } n = 2, 4, 6, \dots. \tag{11.185}$$

As before, n must be a positive, even integer. Equation 11.185 can never be satisfied for any of the acceptable values of n, and so the series for $\grave{B}_1(x)$ never converges for $x = \pm 1$. Equation 11.184, however, is satisfied with $n = 2$. So for $\ell = 1$, the only series that converges for $x = \pm 1$ is the one for $\grave{A}_1(x)$ and we have

$$\grave{P}_1(x) = \grave{A}_1(x) = x. \tag{11.186}$$

This pattern will continue as ℓ is increased. The terminating solutions alternate between the $\grave{A}_\ell(x)$ series and the $\grave{B}_\ell(x)$ series. The terminating solution is always in the form of a polynomial in x, with the order of the polynomial increasing with ℓ.

What about negative values of ℓ? It turns out that negative, integer values of ℓ do not generate new solutions. For example, the $\ell = -1$ solution is identical to the $\ell = 0$ solution, and the $\ell = -2$ solution is the same as the $\ell = 1$ solution. This is not surprising because ℓ appears as the quantity $\ell(\ell + 1)$ in the differential equation. Setting $\ell = -2$ produces the same value of $\ell(\ell + 1)$ as does $\ell = 1$. So to cover all the possible terminating solutions, we only need consider integer values of ℓ greater than or equal to zero.

These terminating solutions are referred to as the Legendre polynomials. The first six are

$$\grave{P}_0(x) = 1$$
$$\grave{P}_1(x) = x$$
$$\grave{P}_2(x) = (1/2)(3x^2 - 1)$$
$$\grave{P}_3(x) = (1/2)(5x^3 - 3x)$$
$$\grave{P}_4(x) = (1/8)(35x^4 - 30x^2 + 3).$$

Each of these solutions has been normalized so that $\grave{P}_\ell(1) = 1$. There is a clever equation, called Rodrigues's formula, for generating these polynomials:

$$\grave{P}_\ell(x) = \frac{1}{2^\ell \ell!} \frac{d^\ell (x^2 - 1)^\ell}{dx^\ell}. \tag{11.187}$$

The general solution to Laplace's equation with $m = 0$ is a linear combination of the $\grave{P}_\ell(\cos \theta)$ and the r-dependent solutions found in Equation 11.163:

$$\Phi(r, \theta) = \sum_{\ell=0}^{\infty} \begin{Bmatrix} r^\ell \\ r^{-\ell-1} \end{Bmatrix} \grave{P}_\ell(\cos \theta). \tag{11.188}$$

Notice, by the arguments given above, only integer values of ℓ greater than or equal to zero appear in the sum.

11.4.2 Orthogonality of the Legendre Polynomials

From our experience with general solutions to Laplace's equation in Cartesian and cylindrical geometries, it is clear that, in order to satisfy arbitrary boundary conditions using an expansion, some of the functions in the expansion must exhibit orthogonality properties. If there are two sums in the general solution, two of the expansion functions must have orthogonality properties. If there is only one summation then only one of the functions needs to have orthogonality properties. In Equation 11.188 there is only one summation and so either the r-dependence functions or the Legendre polynomials must obey an orthogonality condition. It isn't difficult to see that the r-dependent functions will not be able to do this. So we must investigate the Legendre polynomials to determine their orthogonality properties.

Remember the Legendre polynomials were the solutions to the differential equation

$$\frac{\partial}{\partial x}\left[(1-x^2)\frac{\partial}{\partial x}\right]\dot{P}_\ell(x) = -\ell(\ell+1)\dot{P}_\ell(x). \qquad (11.189)$$

The linear operator associated with this equation is

$$\frac{\partial}{\partial x}\left[(1-x^2)\frac{\partial}{\partial x}\right], \qquad \cdot \qquad (11.190)$$

which is in Sturm-Liouville form. We learned earlier that an operator of this type is Hermitian if Equation 11.57 is obeyed. That equation, for this case with $p(x) = 1 - x^2$, becomes

$$\dot{P}_\ell(x)(1-x^2)\frac{d\dot{P}_{\ell'}(x)}{dx}\Bigg|_{x_{min}}^{x_{max}} = 0, \qquad (11.191)$$

where ℓ and ℓ' are two different positive integer values. We defined x_{max} and x_{min} to be the upper and lower limits of the problem. Generally, if the problem includes the z-axis, then the range of θ is $0 \le \theta \le \pi$, and thus $x_{min} = -1$ and $x_{max} = 1$. Consequently, the zeroing of the LHS of Equation 11.191 is accomplished automatically through the $(1 - x^2)$ term. With this definition for Ω and using the Legendre polynomials, the operator of Equation 11.190 is Hermitian.

Equation 11.189 is also in the form of an eigenvalue equation. The eigenvalues are $-(\ell)(\ell+1)$ and the eigenfunctions are the $\dot{P}_\ell(x)$. Since we have already established the Hermitian property, we can immediately write down the orthogonality relation

$$\int_{-1}^{+1} dx\, \dot{P}_\ell(x)\dot{P}_{\ell'}(x) = 0 \qquad \ell \ne \ell'. \qquad (11.192)$$

If we evaluate the above integral when $\ell = \ell'$, the orthogonality condition for these polynomials becomes

$$\int_{-1}^{+1} dx\, \dot{P}_\ell(x)\dot{P}_{\ell'}(x) = \frac{2}{2\ell+1}\delta_{\ell\ell'}. \qquad (11.193)$$

In terms of θ:

$$\int_0^\pi d\theta \, \sin\theta \, \dot{P}_\ell(\cos\theta)\dot{P}_{\ell'}(\cos\theta) = \frac{2}{2\ell+1}\delta_{\ell\ell'}. \qquad (11.194)$$

11.4.3   Solutions Valid off the z-Axis

If a spherical problem excludes the z-axis from consideration, the requirement that the solutions be finite at $\theta = 0$ and $\theta = \pi$ ($x = \pm 1$) can be relaxed. A set of boundary conditions which leads to this type of solution is shown in Figure 11.18. These boundary conditions exist on a conical surface located at $\theta = \theta_o$ and $\theta = (\pi - \theta_o)$. The potential is held at $+V_o$ on the top part of the cone, and $-V_o$ on the bottom. A solution for the potential in the region $\theta_o < \theta < (\pi - \theta_o)$ is desired.

The general solution to Laplace's equation for this type of problem develops along the lines of the previous section. If there is no ϕ-variation, the separable solution can be written as

$$\Phi(r,\theta) = \frac{R(r)}{r}Q(\theta), \qquad (11.195)$$

where $R(r)$ is the same function we developed previously,

$$\frac{R(r)}{r} \rightarrow \begin{cases} r^\ell \\ r^{-\ell-1} \end{cases}, \qquad (11.196)$$

and $Q(\theta)$ is a new solution to Equation 11.158 with $m = 0$, which does not need to be finite on the z-axis. We use the function $Q(\theta)$, because we want to reserve the $P(\theta)$ notation for the Legendre polynomial solutions we derived in the previous section.

The $Q(\theta)$ solutions, of course, depend on the value of ℓ and so we will, as before, use its value as an index: $Q(\theta) \rightarrow Q_\ell(\theta)$. Again we will make the substitution
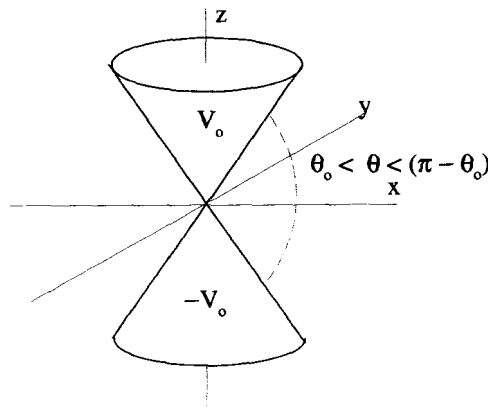
Figure 11.18   Spherical Boundary Conditions that Exclude the z-Axis

$x = \cos\theta$ and $\grave{Q}_\ell(\cos\theta) = Q_\ell(\theta)$. The method of Frobenius can be used to develop the two solutions for $\grave{Q}_\ell(x)$, and they will be identical in form to Equations 11.172 and 11.173. The difference now, however, is that we must include the series solutions that do not terminate, the very ones we removed from the $\grave{P}_\ell(\theta)$ solutions because they did not converge at $x = \pm 1$.

For $\ell = 0$, we will define

$$\grave{Q}_0(x) = \grave{A}_0(x), \tag{11.197}$$

where $\grave{A}_0(x)$ is the Frobenius series of Equation 11.172 which we previously found unacceptable:

$$\grave{Q}_0(x) = x + (1/3)x^3 + (1/5)x^5 + \cdots. \tag{11.198}$$

With some work this series can be expressed in closed form as

$$\grave{Q}_0(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right). \tag{11.199}$$

Similarly for $\ell = 1$, define $\grave{Q}_1(x)$ as the nonterminating series,

$$\grave{Q}_1(x) = \grave{B}_1(x), \tag{11.200}$$

which in closed form can be written as

$$\grave{Q}_1(x) = \frac{x}{2}\ln\left(\frac{1+x}{1-x}\right) - 1. \tag{11.201}$$

For $\ell = 2$:

$$\grave{Q}_2(x) = \frac{3x^2 - 1}{4}\ln\left(\frac{1+x}{1-x}\right) - \frac{3x}{2}. \tag{11.202}$$

There is a generating function for the $\grave{Q}_\ell(x)$ functions that is rather complicated. It involves the Legendre polynomials

$$\grave{Q}_\ell(x) = \frac{1}{2}\grave{P}_\ell(x)\ln\left(\frac{1+x}{1-x}\right) - \frac{2\ell - 1}{\ell}\grave{P}_{\ell-1}(x) - \frac{2\ell - 5}{3(\ell - 1)}\grave{P}_{\ell-3}(x). \tag{11.203}$$

Notice how these solutions diverge for $x = \pm 1$.

Keep in mind, for the $Q_\ell(\theta)$ solutions, nothing says ℓ must be an integer. The particular boundary conditions of the problem being solved will usually limit the possible values of ℓ to some discrete set. Often this is a set of integers. With integer values for ℓ, the infinite series solutions for the $\grave{Q}_\ell(x)$ solutions can be expressed in closed form and paired with the $\grave{P}_\ell(x)$ solutions to give two independent solutions, as summarized in Table 11.1. The $\grave{P}_\ell(x)$ and $\grave{Q}_\ell(x)$ solutions for $\ell = 0, 1,$ and 2 are shown in Figures 11.19 and 11.20.

TABLE 11.1. The $\grave{P}_\ell(x)$ and $\grave{Q}_\ell(x)$ Solutions to Laplace's Equation in Cylindrical Geometry

ℓ	$s = 1$ Solution	$s = 0$ Solution
0	$\grave{Q}_0(x)$	$\grave{P}_0(x)$
1	$\grave{P}_1(x)$	$\grave{Q}_1(x)$
2	$\grave{Q}_2(x)$	$\grave{P}_2(x)$
...	...	....
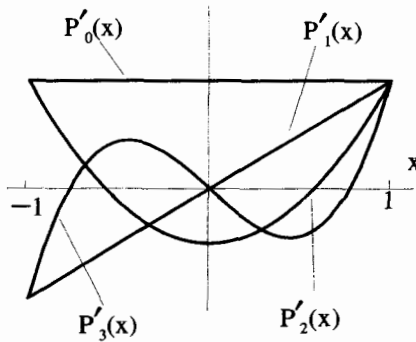
Figure 11.19 The First Four $\grave{P}_\ell(x)$ Functions
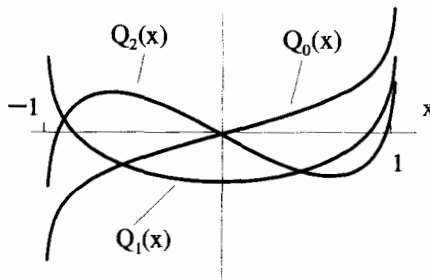
Figure 11.20 The First Three $\grave{Q}_\ell(x)$ Functions

11.4.4 Solutions with ϕ-Dependence

Now we return to the solution of Equation 11.161, this time allowing a ϕ-dependence. Remember the separated solution to Laplace's equation in spherical coordinates had the form

$$\Phi(r, \theta, \phi) = \frac{R(r)}{r}F(\phi)P(\theta),\tag{11.204}$$

where

$$\frac{R(r)}{r} \rightarrow \begin{cases} r^\ell \\ r^{-\ell-1} \end{cases}\tag{11.205}$$

and

$$F(\phi) \rightarrow \begin{cases} e^{+im\phi} \\ e^{-im\phi} \end{cases}.$$ (11.206)

A substitution of $x = \cos\theta$ made $P(\theta) = \grave{P}(\cos\theta)$, and $\grave{P}(x)$ solved the equation

$$\left\{ \frac{\partial}{\partial x} \left[(1 - x^2)\frac{\partial}{\partial x} \right] - \frac{m^2}{1 - x^2} + \ell(\ell + 1) \right\} \grave{P}(x) = 0.$$ (11.207)

From the form of Equation 11.207, it is clear that $\grave{P}(x)$ depends on both m and ℓ, so we will label the solutions of $\grave{P}(x)$ with two indices: $\grave{P}(x) \rightarrow \grave{P}_{\ell m}(x)$.

A brute force Frobenius solution to Equation 11.207 is very messy. It is simplified, somewhat, by making the substitution

$$\grave{P}_{\ell m}(x) = (1 - x^2)^{m/2} \grave{U}_{\ell m}(x).$$ (11.208)

With this, Equation 11.207 becomes

$$\left\{ (1 - x^2) \frac{\partial^2}{\partial x^2} - 2(m + 1)x\frac{\partial}{\partial x} - [m + m^2 - \ell(\ell + 1)] \right\} \grave{U}_{\ell m}(x) = 0.$$ (11.209)

Using the method of Frobenius, we will try to find a solution of the form:

$$\grave{U}_{\ell m}(x) = \sum_{n=0}^{\infty} a_n x^{n+s}.$$ (11.210)

When this is substituted into 11.209, equations for the a_n coefficients are obtained. For $n = 0$ and $n = 1$, the two indicial equations are

$$a_0(s)(s - 1) = 0$$ (11.211)

and

$$a_1(s + 1)(s) = 0.$$ (11.212)

For $n \geq 2$, the generating equation for the a_n's is

$$a_n = a_{n-2} \frac{(n + s - 2)(n + s + 2m - 1) + m + m^2 - \ell(\ell + 1)}{(n + s)(n + s - 1)}.$$ (11.213)

There are two solutions for the indicial equations, $s = 1$ and $s = 0$. Thus we obtain two different solutions,

$$\grave{A}_{\ell m}(x) = \sum_{n=0,2,\dots}^{\infty} a_n x^{n+1} \quad s = 1$$ (11.214)

and

$$\grave{B}_{\ell m}(x) = \sum_{n=0,2,\dots}^{\infty} b_n x^n \qquad s = 0, \qquad (11.215)$$

where, as usual, we have defined $a_0 \neq 0$, $b_0 \neq 0$, and $a_1 = b_1 = 0$. The generating equation for the a_n's, for $n \geq 2$ becomes

$$a_n = a_{n-2} \frac{(n - 1)(n + 2m) + m + m^2 - \ell(\ell + 1)}{(n + 1)(n)} \qquad (11.216)$$

and the one for the b_n's becomes

$$b_n = b_{n-2} \frac{(n - 2)(n + 2m - 1) + m + m^2 - \ell(\ell + 1)}{n(n - 1)}. \qquad (11.217)$$

As $n \to \infty$, both $a_n = a_{n-2}$ and $b_n = b_{n-2}$. This means that for these solutions to be valid on the z-axis (where $x = \pm 1$), the series solutions must terminate. This is the same argument used in the derivation of the Lengendre polynomials.

Let's look at the termination of the $s = 0$ series, the one for $\grave{B}_{\ell m}(x)$. For this series to terminate

$$(n - 2)(n + 2m - 1) + m + m^2 - \ell(\ell + 1) = 0. \qquad (11.218)$$

With some rearrangement this condition for termination becomes

$$(n + m - 2)(n + m - 1) = \ell(\ell + 1). \qquad (11.219)$$

It should be noted that this simple condition for termination was what originally motivated the substitution of Equation 11.208. Had we not made that substitution, this condition would have been much more difficult to deal with! At this point, all we know about the terms in Equation 11.219 is that, according to Equation 11.150, m must be a nonnegative integer, and that n is a positive, even integer. This means that the first factor on the LHS of Equation 11.219, $(n + m - 2)$, is an integer and that the second factor, $(n + m - 1)$, is equal to that same integer plus one. Therefore, if Equation 11.219 is to be satisfied, and the series is to terminate, ℓ must be an integer. As was the case before, the solutions derived from considering negative values of ℓ are just repetitions of the positive ℓ solutions. Therefore, we need only consider $\ell = 0, 1, 2, \dots$.

An interesting result of Equation 11.219 is a new condition on m. Pick some positive, integer value for ℓ, say $\ell = \ell_o$. Termination then occurs when

$$(n + m - 2)(n + m - 1) = \ell_o(\ell_o + 1). \qquad (11.220)$$

If we set $m = \ell_o$, termination can occur with $n = 2$. If $m > \ell_0$ the LHS of Equation 11.220 is always greater than the RHS and termination can never occur. When $m = \ell_0 - 1$, there is no way to satisfy Equation 11.220 with an even value of n.

If we move m down another notch, and set $m = \ell_0 - 2$, we can satisfy Equation 11.220 again with a choice of $n = 4$. If we continue this process, we will find we can satisfy the condition in general for every other integer value of m until we reach either $m = 1$ or $m = 0$, depending on whether ℓ_o was even or odd. That is, the values of m can be

$$m = \begin{cases} \ell_o & (n = 2) \\ \ell_o - 2 & (n = 4) \\ \ell_o - 4 & (n = 6) \\ \vdots & \vdots \\ 1 \text{ or } 0 \end{cases} \tag{11.221}$$

Remember this requirement is for the $s = 0$ series.

If we determine, in a similar way, the termination condition for the $s = 1$ series, we find the relationship

$$(n + m - 1)(n + m) = \ell_o(\ell_o + 1). \tag{11.222}$$

The highest value that m can take on and still satisfy this equation is $\ell_o - 1$, with the $n = 2$. If m gets higher than this value, it is impossible to satisfy the relationship. If we try $m = \ell_o - 2$, again we find we cannot satisfy Equation 11.222. But when we move down to $m = \ell_o - 3$, we can satisfy it with $n = 4$. Much like we derived Equation 11.221, we find all the possible m values are given by

$$m = \begin{cases} \ell_o - 1 & (n = 2) \\ \ell_o - 3 & (n = 4) \\ \ell_o - 5 & (n = 6) \\ \vdots & \vdots \\ 1 \text{ or } 0 \end{cases} \tag{11.223}$$

where again, the lowest value of m depends on the starting value of ℓ_o.

When we consider both the $s = 1$ and $s = 0$ series solutions collectively, we find an important fact. To have at least one valid solution, m must obey either Equation 11.221 or Equation 11.223. In other words, $0 \leq m \leq \ell_o$.

Now that we know the range of m, we need to determine what the terminating $\grave{P}_{\ell m}(x)$ solutions actually look like. We will start with the $\ell = 0$ solution. According to the previous discussion, with $\ell = 0$ there can only be a terminating series solution if $m = 0$, and it must come from the $s = 0$ series with $n = 2$. Therefore, using Equation 11.208,

$$\grave{P}_{00}(x) = \grave{U}_{00}(x)$$
$$= \grave{B}_{00} \tag{11.224}$$
$$= b_o.$$

The particular value of b_o is usually chosen as 1. This solution is not terribly exciting, since it is just a constant.

With $\ell = 1$, there are two possible m values: $m = 1$ and $m = 0$. The solution $\grave{P}_{11}(x)$ uses the $s = 0$ series with $n = 2$ so

$$
\begin{aligned}
\grave{P}_{11}(x) &= (1 - x^2)^{1/2}\grave{U}_{11}(x) \\
&= (1 - x^2)^{1/2}\grave{B}_{11}(x) \\
&= (1 - x^2)^{1/2}b_0.
\end{aligned}
\tag{11.225}
$$

As before, we usually set $b_o = 1$. The $\grave{P}_{10}(x)$ solution uses the $s = 1$ series with $n = 2$, which gives

$$
\begin{aligned}
\grave{P}_{10}(x) &= \grave{U}_{10}(x) \\
&= \grave{A}_{10}(x) \\
&= a_0 x.
\end{aligned}
\tag{11.226}
$$

If we set $a_0 = 1$, $\grave{P}_{10}$ equals the $\grave{P}_1(x)$ Legendre polynomial we derived earlier.

In this way the solutions for all the terminating $\grave{P}_{\ell m}(x)$ solutions, valid on the z-axis, can be developed. The solutions are collectively referred to as the associated Legendre polynomials. A generating function for these polynomials, a generalization of Rodrigues's formula, has been developed:

$$
\grave{P}_{\ell m}(x) = \frac{1}{2^\ell \ell!}(1 - x^2)^{m/2}\frac{\partial^{\ell+m}(x^2 - 1)^\ell}{\partial x^{\ell+m}}.
\tag{11.227}
$$

These functions obey the orthogonality condition

$$
\int_{-1}^1 dx\, \grave{P}_{\ell m}(x)\grave{P}_{\ell'm}(x) = \left(\frac{2}{2\ell + 1}\right)\frac{(\ell + m)!}{(\ell - m)!}\delta_{\ell\ell'},
\tag{11.228}
$$

which in terms of the variable θ is

$$
\int_0^\pi d\theta\, \sin\theta\grave{P}_{\ell m}(\cos\theta)\grave{P}_{\ell'm}(\cos\theta) = \left(\frac{2}{2\ell + 1}\right)\frac{(\ell + m)!}{(\ell - m)!}\delta_{\ell\ell'}.
\tag{11.229}
$$

Finally, the general solution to Laplace's equation in spherical coordinates using the $R(r)$, $F(\phi)$, and $P(\theta)$ functions can be written down. The result is

$$
\Phi(r, \theta, \phi) = \sum_{\ell=0}^\infty \sum_{m=0}^{+\ell} \begin{Bmatrix} r^\ell \\ r^{-\ell-1} \end{Bmatrix} \begin{Bmatrix} e^{+im\phi} \\ e^{-im\phi} \end{Bmatrix} \grave{P}_{\ell m}(\cos\theta).
\tag{11.230}
$$

Example 11.6 We will now use these results to find the electric potential inside the spherical surface shown in Figure 11.21. Suppose that different electric potentials have been applied on each quadrant of the surface, as indicated in the figure. We will only show a general method of solution here and leave complete solutions for the exercises at the end of this chapter.

$$\Phi(R_o, \theta, \phi) = \begin{array}{ll} +V_o & 0 < \theta < \pi/2, \ \ 0 < \phi < \pi \\ -V_o & 0 < \theta < \pi/2, \ \ \pi < \phi < 2\pi \\ +V_o & \pi/2 < \theta < \pi, \ \ \pi < \phi < 2\pi \\ +V_o & \pi/2 < \theta < \pi, \ \ 0 < \phi < \pi \end{array}$$
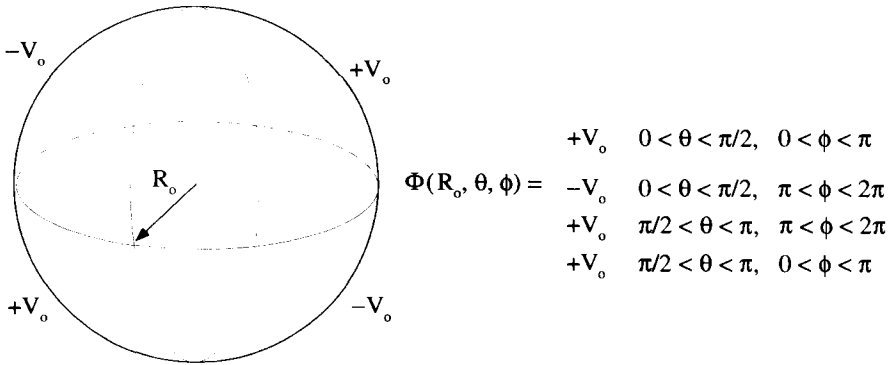
Figure 11.21 Boundary Conditions on a Spherical Surface

Start with the general spherical solution in the form

$$\Phi(r, \theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=0}^{\ell} \left\{ \begin{array}{l} r^{\ell} \\ r^{-\ell-1} \end{array} \right\} \left\{ \begin{array}{l} \cos(m\phi) \\ \sin(m\phi) \end{array} \right\} \grave{P}_{\ell m}(\cos\theta). \qquad (11.231)$$

We have chosen to use trigonometric functions for the ϕ-dependence instead of complex exponentials, because the electrostatic potential should be a real quantity.

Immediately, we can make some useful simplifications. All the $r^{-\ell-1}$ solutions must vanish because the potential must be finite at $r = 0$. In addition, the $\cos(m\phi)$ solutions are also zero because the boundary conditions, and therefore the solution, have odd symmetry about ϕ. This reduces Equation 11.231 to

$$\Phi(r, \theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=0}^{\ell} A_{\ell m} r^{\ell} \sin(m\phi)\grave{P}_{\ell m}(\cos\theta). \qquad (11.232)$$

Notice that there is a single amplitude constant $A_{\ell m}$, which is indexed by both ℓ and m.

Values for this constant are determined by requiring the solution to go to the prescribed values at the boundary $r = R_o$:

$$\Phi(R_o, \theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=0}^{\ell} A_{\ell m} R_o^{\ell} \sin(m\phi)\grave{P}_{\ell m}(\cos\theta). \qquad (11.233)$$

A particular m value can be selected using the orthogonality of the sine functions. A particular ℓ value can be isolated using the orthogonality of the associated Legendre polynomials. Using these orthogonality conditions allows a solution for $A_{\ell m}$:

$$A_{\ell m} = \left(\frac{2\ell + 1}{2\pi R_o^{\ell}} \right) \frac{(\ell - m)!}{(\ell + m)!} \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \ \sin\theta \sin(m\phi)\grave{P}_{\ell m}(\cos\theta)\Phi(R_o, \theta, \phi).$$

$$(11.234)$$

This integral determines the $A_{\ell m}$, which are then inserted into Equation 11.232 for the solution inside the spherical surface.

11.4.5   Spherical Harmonics

It is customary to combine the $e^{im\phi}$ functions with the associated Legendre polynomials in Equation 11.230 to define a new function of ϕ and θ, the so-called spherical harmonics:

$$\underline{Y}_{\ell m}(\theta, \phi) = (-1)^m \sqrt{\frac{(2\ell + 1)(\ell - m)!}{4\pi(\ell + m)!}} e^{im\phi} \grave{P}_{\ell m}(\cos\theta). \tag{11.235}$$

The complicated multiplying factor just normalizes the function, so that the orthogonality condition, discussed shortly, is as simple as possible. The $(-1)^m$ factor is a convention used frequently in quantum mechanics. Notice, because we separated out the $e^{im\phi}$ factor, m must now be allowed to vary between $-\ell$ and $+\ell$. To make this work, we define the associated Legendre polynomials with negative values of m to simply be proportional to the same polynomial with a positive m:

$$\grave{P}_{\ell,-m}(x) = \frac{(-1)^m(\ell - m)!}{(\ell + m)!}\grave{P}_{\ell,m}(x). \tag{11.236}$$

The weird proportionality constant makes Equation 11.227 continue to hold for all values of m.

Using the spherical harmonics, the general solution of Equation 11.230 becomes

$$\Phi(r, \theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{+\ell} \left\{ \begin{matrix} r^\ell \\ r^{-\ell-1} \end{matrix} \right\} \underline{Y}_{\ell m}(\theta, \phi). \tag{11.237}$$

The spherical harmonics obey an orthogonality condition:

$$\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \, \sin\theta \, \underline{Y}_{\ell m}^*(\theta, \phi)\underline{Y}_{\ell'm'}(\theta, \phi) = \delta_{\ell\ell'}\delta_{mm'}. \tag{11.238}$$

Notice, since the spherical harmonics are complex functions, the orthogonality relation involves a complex conjugate operation.

The first few $Y_{\ell m}$'s are:

$$\underline{Y}_{00}(\theta, \phi) = 1/\sqrt{4\pi}$$

$$\underline{Y}_{1-1}(\theta, \phi) = \sqrt{3/8\pi}\sin\theta e^{-i\phi}$$

$$\underline{Y}_{10}(\theta, \phi) = \sqrt{3/4\pi}\cos\theta$$

$$\underline{Y}_{11}(\theta, \phi) = -\sqrt{3/8\pi}\sin\theta e^{i\phi}$$

$$\underline{Y}_{2-2}(\theta, \phi) = \sqrt{45/96\pi} \sin^2\theta e^{-i2\phi} \qquad (11.239)$$

$$\underline{Y}_{2-1}(\theta, \phi) = \sqrt{45/24\pi} \sin\theta \cos\theta e^{-i\phi}$$

$$\underline{Y}_{20}(\theta, \phi) = \sqrt{5/4\pi}\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right)$$

$$\underline{Y}_{21}(\theta, \phi) = -\sqrt{45/24\pi} \sin\theta \cos\theta e^{i\phi}$$

$$\underline{Y}_{22}(\theta, \phi) = \sqrt{45/96\pi} \sin^2\theta e^{i2\phi}.$$

These functions can be visualized by making three-dimensional, polar plots of $|\underline{Y}_{\ell m}(\theta, \phi)|^2$ vs. θ and ϕ. The plot for $\underline{Y}_{00}(\theta, \phi)$ is shown in Figure 11.22, where the magnitude squared of the spherical harmonic is represented by the distance from the origin to the surface. Similar surfaces are shown for the $\ell = 1$ and $\ell = 2$ spherical harmonics in Figures 11.23 and 11.24. In these figures, the z-axis is vertical, and the surfaces are always symmetric about this axis because $|e^{im\phi}|^2 = 1$. Also, the plots for $\pm m$ are always exactly the same, because $|Y_{\ell m}| = |Y_{\ell -m}|$.

The relevance of these spherical harmonics to the quantum mechanical wave functions for the hydrogen atom should be clear to anyone with a chemistry or physics background. It is important to realize that the restriction on m, i.e., that $|m| \leq \ell$, is a mathematical result, originating from our requirement that the solutions converge for all values of θ.

$$\underline{Y}_{00}(\theta,\phi)$$
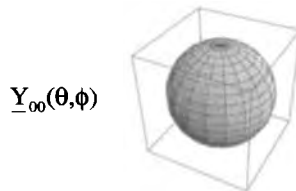
Figure 11.22 Surface Representing $\underline{Y}_{00}(\theta, \phi)$

$$\underline{Y}_{11}(\theta,\phi) \qquad\qquad \underline{Y}_{10}(\theta,\phi) \qquad\qquad \underline{Y}_{1-1}(\theta,\phi)$$
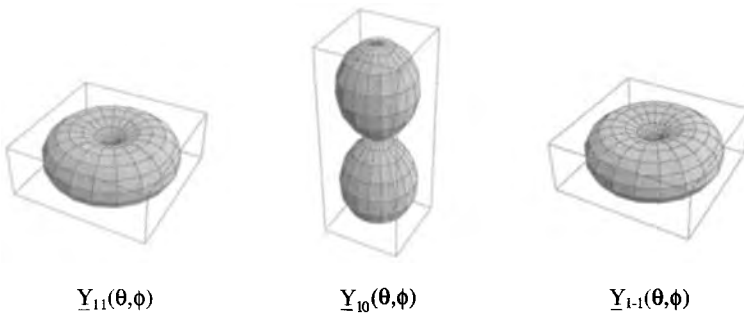
Figure 11.23 Surfaces Representing the $\underline{Y}_{1m}(\theta, \phi)$

$\underline{Y}_{22}(\theta,\phi)$ $\underline{Y}_{21}(\theta,\phi)$

$\underline{Y}_{2\text{-}2}(\theta,\phi)$ $\underline{Y}_{2\text{-}1}(\theta,\phi)$ $\underline{Y}_{20}(\theta,\phi)$
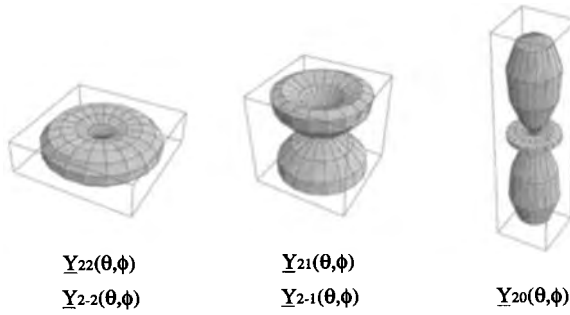
Figure 11.24   Surfaces Representing the $\underline{Y}_{2m}(\theta,\phi)$

EXERCISES FOR CHAPTER 11

1. Find a series solution for the electric potential inside a two-dimensional square box, with sides held at the potentials shown below:

$\Phi = 0$

L $\Phi = +V_o$ $\Phi = +V_o$

$\Phi = 0$
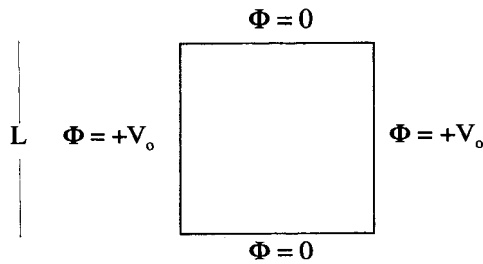
2. Find the two-dimensional solution to Laplace's equation inside a rectangular box of length a and height b, if the boundary conditions for the solution are as shown in the figure below:

2

b   1   1

0

a
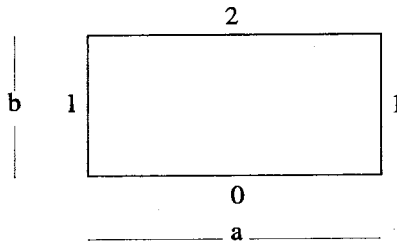
3. Find the solution to Laplace's equation inside a 1×1, two-dimensional square region, if the solution on the boundary is as described in the figure below:

y

$\phi = x$

$\phi = 0$ $\phi = y$

x

$\phi = 0$
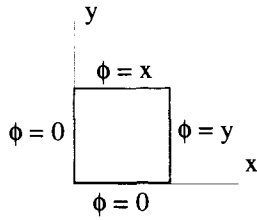
4. Consider the $L \times 3L$ rectangular box shown below.

      0   +1   0

L   0   0

      0   +1   0

3L
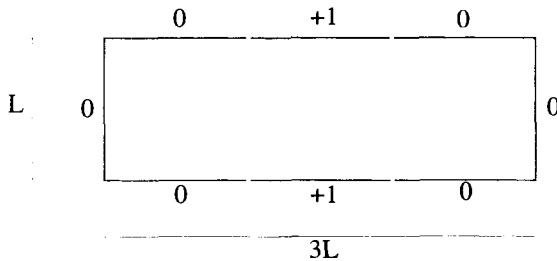
The electrostatic potential on the boundaries of this box is indicated in this figure. The end caps of the box are held at $\phi = 0$ zero, while the central section has $\Phi = 1$. What is the potential everywhere inside the box?

5. Find a series solution for the potential inside the two-dimensional rectangular box, with sides held at $\pm V_o$, as shown in the figure below:

$\Phi = -V_o$

L_o $\Phi = +V_o$ $\Phi = +V_o$

$\Phi = -V_o$

$2L_o$
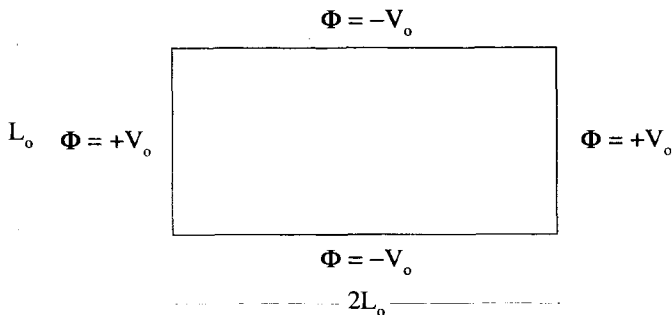
6. Find $\Phi(x, y)$, the solution to Laplace's equation in two dimensions between the two parallel planes $y = 0$ and $y = a_o$. Assume $\Phi(x, 0) = 0$, and that $\Phi(x, a_o)$ is periodic with period $2b_o$. Over one period, this function is given by

$$\Phi(x, a_o) = \begin{cases} +1 & 0 < x < b_o \\ -1 & b_o < x < 2b_o \end{cases},$$

and is periodic with period $2b_o$ for all x.

7. Find the two-dimensional solutions to Laplace's equation inside the two isosceles right triangles, with the boundary conditions shown in the figures below:
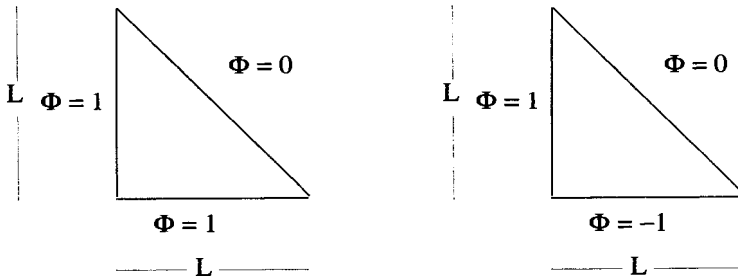
8. Consider the following two-dimensional boundary conditions and the solution to Laplace's equation inside this boundary.
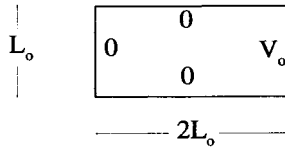
(a) Set up the periodic boundary conditions that will generate a solution to Laplace's equation inside this rectangular region and indicate the region of this periodic problem where its solution is equal to the solution of this rectangular problem.

(b) Obtain a solution to Laplace's equation inside the rectangular region that satisfies the above boundary conditions.

(c) Repeat the above steps for the boundary conditions shown in the following figure.
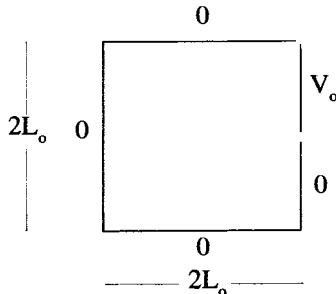
9. For this problem, you are to develop the Green's function for a square drum head lying in the xy-plane, as shown below. The boundaries at $x = 0, L$ and $y = 0, L$ are held at $h = 0$.
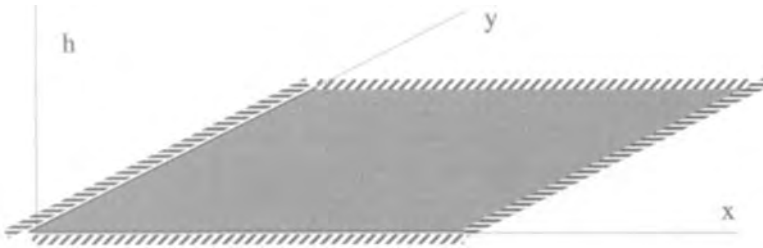
The displacement of the vibrating drum head $h(x, y)e^{i\omega t}$ is governed by the partial differential equation

$$\nabla^2 h(x, y) + \frac{\omega^2}{c^2} h(x, y) = d(x, y),$$

where $d(x, y)e^{i\omega t}$ is a driving function. The Green's function for this problem therefore satisfies

$$\left(\nabla^2 + \frac{\omega^2}{c^2}\right) g(x|\xi_x, y|\xi_y) = \delta(x - \xi_x)\delta(y - \xi_y).$$

(a) As the first step in finding the Green's function, turn this problem into one that is infinite in the x- and y-directions. Indicate the locations and nature of the driving δ-functions for a periodic problem whose solution is identical to the drum head problem in the region $0 < x < L$ and $0 < y < L$.

(b) Find a series solution for $g(x|\xi_x, y|\xi_y)$.

10. The potential between two concentric, electrically conducting cylindrical surfaces is a simple electrostatic problem. If the inside cylinder of radius r_1 is held at a potential V_1, and the outer cylinder at r_2 is held at V_2, the potential between the surfaces is

$$\Phi(r) = \frac{V_1 - V_2}{\ln(r_1/r_2)} \ln \rho + \frac{V_1 \ln r_2 - V_2 \ln r_1}{\ln(r_2/r_1)}.$$

Verify this expression and show how it is given by the general cylindrical solution presented in this chapter.

11. Consider the two-dimensional polar form of Laplace's equation

$$\left[\frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2}{\partial \theta^2}\right] u(\rho, \theta) = 0$$

and the separated solutions of the form $u(\rho, \theta) = R(\rho)H(\theta)$.

(a) What are solutions for $H(\theta)$?

(b) Determine the two independent solutions for $R(\rho)$.

(c) Using the results from parts (a) and (b), determine the two-dimensional solution to Laplace's equation inside a circle of radius r_o, if on the circle, the solution goes to $u(r_o, \theta) = V_o \sin \theta$. What is the potential outside this circle, for $\rho > r_o$? What assumptions have to be made to get the answer for $\rho > r_o$?

12. Find the solution to Laplace's equation inside a cylinder of radius r_o and length L_o with the boundary conditions shown in the figure below:
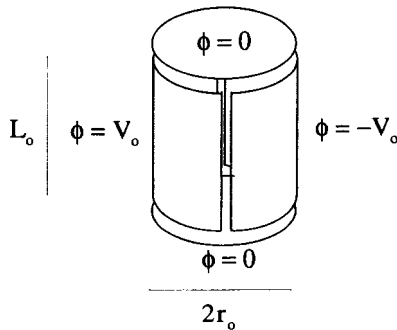
Now find the solution inside the same cylinder with the boundary conditions defined as shown in the figure below:
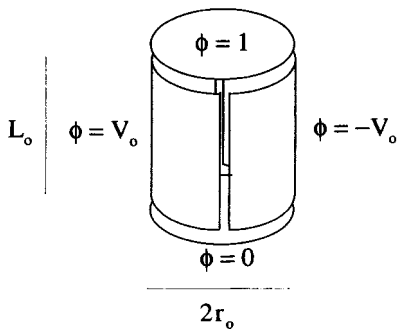
13. Find the solution to Laplace's equation inside a cylinder of radius r_o and length L_o with the following boundary conditions: The sides and bottom are held at $\Phi = 0$ and the top is held at $\Phi = V_o(1 - \rho/r_o)$, where $\rho = 0$ is the axis of the cylinder. Use a reference such as Abramowitz and Stegun to at least make a stab at the integrals.

14. Find the solution to Laplace's equation inside the hollow cylindrical region of length L between r_1 and r_2, if the solution is to go to zero on the top and bottom and to $+V_o$ and $-V_o$ on the outer and inner surfaces, as shown below:

$$\Phi(r_1) = +V_o$$

$$\Phi(r_2) = -V_o$$
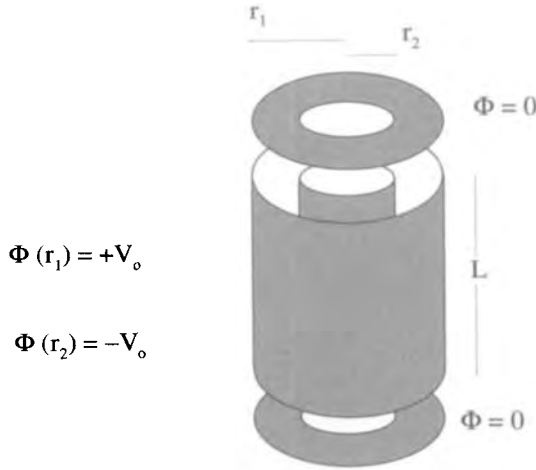
To solve this problem you must construct a new eigenfunction $M_0(k\rho)$, which is a linear combination of $I_0(k\rho)$ and $K_0(k\rho)$, such that $M_0(kr_1) = +1$ and $M_0(kr_2) = -1$.

15. Find the solution to Laplace's equation for the same cylindrical region as described in the previous problem, but with the different boundary conditions shown below:

$$\Phi(r_1) = 0$$

$$\Phi(r_2) = 0$$
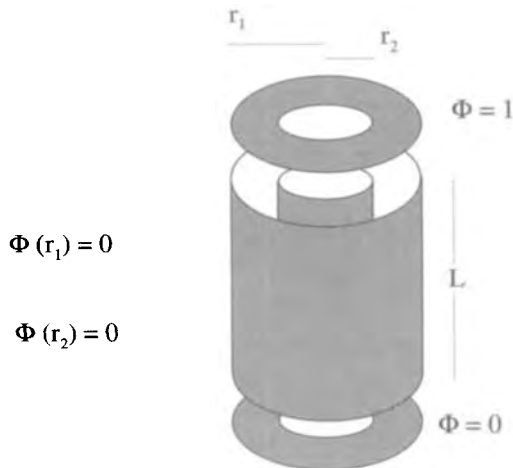
16. For small, static displacements, the deflection of a drum head satisfies the following two-dimensional differential equation in polar coordinates

$$T\nabla^2 z(\rho, \theta) = F(\rho, \theta),$$

where T is the tension in the drum membrane, z is the displacement, and F is the force per unit area applied to the membrane. Let the drum head have a radius r_o, as shown in the figure below.
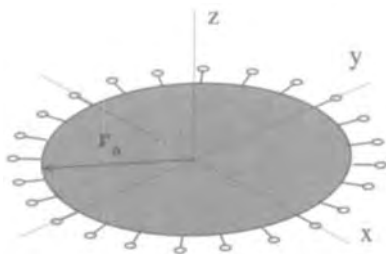
Take the applied force per unit area to be independent of θ, so that the differential equation becomes

$$T\frac{1}{\rho}\frac{d}{d\rho}\rho\frac{dz(\rho)}{d\rho} = F(\rho).$$

(a) Determine the Green's function $g(\rho|\xi)$ that allows the displacement to be written as

$$z(\rho) = \int_o^{r_o} d\xi\, f(\xi)g(\rho|\xi).$$

(b) Plot $g(\rho|\xi)$ vs ρ.

(c) If $F(\rho) = 1 - \rho^2/r_o^2$, set up the Green's function integration for $z(\rho)$.

17. This problem considers the two-dimensional solution to Laplace's equation in cylindrical coordinates from a Green's function point of view. For no z-dependence, find the Green's function $g(\rho, \theta|\theta_o)$ that allows the solution

$$\Phi(\rho, \theta) = \int_0^{2\pi} d\theta_o\, f(\theta_o)g(\rho, \theta|\theta_o).$$

The function $f(\theta) = \Phi(r_o, \theta)$ defines the boundary condition at $\rho = r_o$. Inside this boundary $\Phi(\rho, \theta)$ satisfies a two-dimensional Laplace's equation in cylindrical coordinates

$$\frac{\partial^2\Phi(\rho, \theta)}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial\Phi(\rho, \theta)}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2\Phi(\rho, \theta)}{\partial\theta^2} = 0.$$

The standard way to solve this problem sets $\Phi(\rho, \theta)$ up as a series which must be shown to converge.

18. Using similar steps to those in the previous problem, find the Green's function for the two-dimensional homogeneous equation

$$\nabla^2 u(\rho, \theta) + k^2 u(\rho, \theta) = 0$$

inside a circle of radius $\rho = R_o$, with the boundary condition that $u(R_o, \theta) = f_o(\theta)$.

19. The general form for the solution to Laplace's equation in cylindrical coordinates that is not periodic in z was expressed in this chapter using a shorthand notation that did not include the amplitude constants:

$$\Phi(\rho, \theta, z) = \begin{Bmatrix} J_\nu(k\rho) \\ Y_\nu(k\rho) \end{Bmatrix} \begin{Bmatrix} \sin(\nu\theta) \\ \cos(\nu\theta) \end{Bmatrix} \begin{Bmatrix} \sinh(kz) \\ \cosh(kz) \end{Bmatrix}.$$

If this expression is expanded out, eight amplitude constants need to be introduced. If a solution is sought for $\rho < r_o$, the $Y_\nu(k\rho)$ solutions can be eliminated because they diverge as $\rho \to 0$. Therefore, the number of amplitude constants is reduced to four. If, however we exclude the origin from the domain of the solution, i.e., if we limit the problem to a range $a < \rho < b$, both the $J_\nu(k\rho)$ and $Y_\nu(k\rho)$ functions are necessary and all eight amplitude constants must be considered. The amplitude constants are not the only complication for this $a < \rho < b$ situation, because the spatial factor k is also determined by boundary conditions.

Consider a solution valid for $a < \rho < b$ that is zero at $\rho = a$ and $\rho = b$. To solve this problem, construct a new eigenfunction from the linear sum of the Bessel functions. Call this new eigenfunction $R_\nu(k\rho)$ where

$$R_\nu(k\rho) = A\, J_\nu(k\rho) + B\, Y_\nu(k\rho).$$

Discuss how you would determine the set of values for k and the linear scaling factors A and B, so that this new eigenfunction is normalized and satisfies the zero boundary conditions at $\rho = a$ and $\rho = b$.

20. The potential between two concentric, electrically conducting spherical surfaces is a simple electrostatic problem. If the inside sphere of radius r_1 is held at a potential V_1, and the outer sphere at r_2 is held at V_2, the potential between the surfaces is

$$\Phi(r) = \frac{V_1 r_1 - r_2 V_2}{r_1 - r_2} + \frac{r_1 r_2 (V_1 - V_2)}{r_2 - r_1} \frac{1}{r}.$$

Verify this expression and show how it is given by the general spherical solution presented in this chapter.

21. Find the potential between two concentric spherical surfaces of radius R_o and $2R_o$. The voltage on the upper hemisphere of the outer surface is held at $+V_o$,

and the voltage on the lower hemisphere of the outer surface is held at $-V_o$. The voltage on the upper hemisphere of the inner surface is held at $-V_o$, and the voltage on the lower hemisphere of the inner surface is held at $+.V_o$.

22. A hemispherical shell of radius R_o, held at $\Phi = +V_o$ is positioned on top of an infinite conducting ground plane held at $\Phi = 0$. Use symmetry, in this case sometimes referred to as the method of images, to find the solution to Laplace's equation both inside and outside the hemisphere that satisfies these boundary conditions.

23. Find a series solution for the potential inside a spherical shell whose Hemispheres are held at $+2V_o$ and $-V_o$ as shown below.
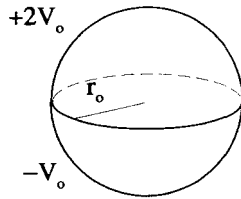
Can you think of an easier way to solve this problem, using the result of the previous exercise?

24. Helmholtz's equation is written as

$$\nabla^2 \, \Psi(\overline{r}) + k^2 \, \Psi(\overline{r}) = 0$$

where k^2 is a positive constant. Show that in spherical coordinates this equation separates and the differential equation the radially dependent component satisfies becomes

$$r^2 \frac{d^2 R_n(r)}{dr^2} + 2r \frac{dR_n(r)}{dr} + \left[k^2 r^2 - n(n + 1)\right] R_n(r) = 0$$

where n is an integer.

(a) Compare this to Bessel's equation.

(b) Show that $R_n(r)$ for different values of n can be combined with the operator and appropriate boundary conditions to satisfy the Hermitian condition. Identify those boundary conditions and the space over which the integration must occur to establish the orthogonality condition for these functions.

(c) Show that the substitution

$$R_n(kr) = \frac{Z_n(kr)}{\sqrt{kr}}$$

gives a form of Bessel's equation, with $Z_n(kr)$ being the Bessel function of order $n + 1/2$. The functions $R_n(kr)$ are called *Spherical Bessel Functions*.

25. The quantum mechanical operators L_+ and L_- are given by

$$L_+ = L_x + iL_y = e^{i\phi}\left[\frac{\partial}{\partial\theta} + i\cot\theta\frac{\partial}{\partial\phi}\right]$$

and

$$L_- = L_x - iL_y = -e^{-i\phi}\left[\frac{\partial}{\partial\theta} - i\cot\theta\frac{\partial}{\partial\phi}\right].$$

The spherical harmonics $Y_{\ell m}(\theta, \phi)$ are not eigenfunctions of these operators. However, interesting things happen when these operators work on the $Y_{\ell m}(\theta, \phi)$. Show that

$$L_+ Y_{\ell m}(\theta, \phi) = \sqrt{(\ell - m)(\ell + m + 1)}Y_{\ell m+1}(\theta, \phi)$$

and

$$L_- Y_{\ell m}(\theta, \phi) = \sqrt{(\ell + m)(\ell - m + 1)}Y_{\ell m-1}(\theta, \phi).$$

Because of this property, L_+ and L_- are called raising and lowering operators.

26. Find the potential between $60° < \theta < 120°$ of two infinite 60° cones, one held at $+2V_o$ and the other at $+V_o$, as shown below:
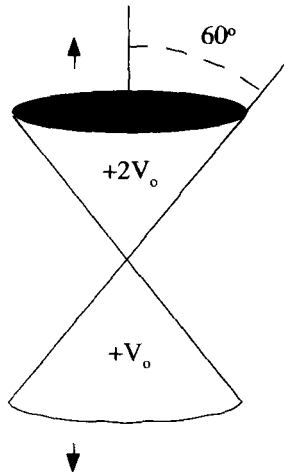
27. Show that the eigenvalues of an Hermitian matrix are pure real quantities and that the eigenvectors are orthogonal.

28. Because of their orthogonality and completeness properties the Legendre polynomials can be used to make series expansions of functions

$$f(x) = \sum_{\ell=0,1,...}^{\infty} C_\ell \grave{P}_\ell(x).$$

Let $f(x)$ be the triangular function over the range $-1 < x < 1$ shown below:
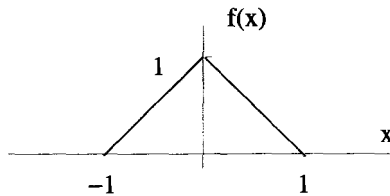
(a) What are the first three coefficients of the Legendre polynomial expansion of this function?

(b) What does the Legendre expansion for this function look like for $|x| > 1$?

(c) Consider a new function which is a similar triangular function, except now it extends from $x = -2$ to $x = 2$. Construct a Legendre polynomial expansion for this function and identify its first three coefficients.

(d) If now $f(x) = 1 - x^2$ over the interval $-1 < x < +1$, what are the first three coefficients of its Legendre polynomial expansion?

29. The Legendre polynomials are determined to within arbitrary constants by the homogeneous differential equations they satisfy:

$$\grave{P}_0(x) = A_0$$
$$\grave{P}_1(x) = A_1 x$$
$$\grave{P}_2(x) = A_2(1 - 3x^2)$$

$$\vdots$$

Determine the constants A_0, A_1, and A_2 for the first three Legendre polynomials, so that the functions obey the orthogonality condition

$$\int_{-1}^{1} dx \, \grave{P}_\ell(x) \grave{P}_{\ell'}(x) = \frac{2}{2\ell + 1} \delta_{\ell\ell'}.$$

30. The following differential equation is set up in eigenvalue/eigenfunction form,

$$\left[\frac{d^2}{d\theta^2} + \frac{\cos\theta}{\sin\theta} \frac{d}{d\theta} \right] P_\lambda(\theta) = -\lambda P_\lambda(\theta),$$

where θ ranges from zero to π, and $P_\lambda(\theta)$ is an eigenfunction of the differential operator with the eigenvalue λ. This operator is not Hermitian.

(a) Place this differential equation in Sturm-Liouville form and show that the operator that results is Hermitian.

(b) Using this new Hermitian operator show that if $m \neq n$

$$\int_0^\pi d\theta \, g(\theta) P_m(\theta) P_n(\theta) = 0$$

and determine the weighting function $g(\theta)$.

31. Consider the differential equation

$$\left(\frac{d^2}{dx^2} + \alpha \frac{d}{dx} \right) \phi_n(x) = -\lambda_n \, \phi_n(x),$$

where α is a positive constant. This equation is in eigenfunction, eigenvalue form.

(a) If λ_n is assumed to be positive and real, show that the two solutions for the eigenfunctions become

$$\phi_n(x) \rightarrow \begin{cases} e^{-\alpha x/2} \sin\left(x\sqrt{\lambda_n - \alpha^2/4} \right) \\ e^{-\alpha x/2} \cos\left(x\sqrt{\lambda_n - \alpha^2/4} \right) \end{cases}$$

and that this form is valid for any value of $\lambda_n \neq \alpha^2/4$.

(b) Now require that the eigenfunctions of part (a) satisfy the following boundary conditions: $\phi_n(0) = 0$ and $\phi_n(L) = 0$. These boundary conditions determine the acceptable values for the λ_n, as well as the acceptable $\phi_n(x)$. Use the n-index to identify these acceptable values for the λ_n and make a plot of the first three acceptable eigenfunctions.

(c) The differential equation given above is not in Sturm-Liouville form. Find the appropriate multiplying function and place it in S-L form.

(d) Determine the orthogonality conditions for the eigenfunctions of part (a) that satisfy the boundary conditions of part (b). Be sure to identify the region over which the functions are orthogonal and explain how you arrived at your answer.

32. The Gram-Schmidt orthogonalization procedure generates a set of orthonormal functions (or vectors) from a set of independent functions (or vectors). For this problem start with the set of independent functions $y_n(x) = x^n$. A number of different orthonormal sets can be constructed from these $y_n(x)$, depending on the weighting function used. Let $\phi_n(x)$ be the set of functions generated from different sums of the $y_n(x)$, and let the orthonormalization be given by the

following integration

$$\int_0^\infty dx\ \phi_n(x)\phi_m(x)e^{-x} = \delta_{nm},$$

where a weighting function $w(x) = e^{-x}$ has been used.

(a) The first function $\phi_0(x)$ is determined by setting it equal to a constant times $y_0(x)$ and then normalizing:

$$\phi_0(x) = a_{00}\ y_0(x).$$

What is $\phi_0(x)$?

(b) The second function $\phi_1(x)$ is set up as a linear sum of $y_1(x)$ and $\phi_0(x)$:

$$\phi_1(x) = a_{11}\ y_1(x) + b_{10}\ \phi_0(x).$$

Find the constants a_{11} and b_{10} that normalize $\phi_1(x)$ and simultaneously make it orthogonal to $\phi_0(x)$.

(c) Continue this process and construct $\phi_2(x)$. Set up $\phi_2(x)$ as a linear sum of $y_2(x)$, $\phi_0(x)$, and $\phi_1(x)$:

$$\phi_2(x) = a_{22}\ y_2(x) + b_{20}\ \phi_0(x) + b_{21}\ \phi_1(x).$$

Find the constants a_{22}, b_{20}, and b_{21} that normalize $\phi_2(x)$ and make it orthogonal to $\phi_0(x)$ and $\phi_1(x)$.

These $\phi_m(x)$ functions are known as Leguerre polynomials. It is interesting to note the Legendre polynomials can be generated by the Gram-Schmidt process using the same set of independent functions $y_n(x) = x^n$ and the orthonormalization given by

$$\int_{-1}^1 dx\ \phi_n(x)\phi_m(x) = \frac{2}{2n+1}\delta_{nm}.$$

33. The elliptical coordinates (u, v, z) are related to a set of Cartesian coordinates by the equations

$$x = \rho_o \cosh u \cos v$$

$$y = \rho_o \sinh u \sin v$$

$$z = z,$$

where ρ_o is a positive constant needed to make the above equations dimensionally correct. Laplace's equation for $\Psi(u, v, z)$ becomes

$$\frac{\partial^2 \Psi(u,v,z)}{\partial z^2} + \frac{1}{2\rho_o^2[\cosh(2u) - \cos(2v)]}\left(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}\right)\Psi(u,v,z) = 0.$$

Assume $\Psi(u, v, z)$ can be separated, i.e., that $\Psi(u, v, z) = f(u)g(v)h(z)$, and determine the three ordinary differential equations that $f(u)$, $g(v)$, and $h(z)$ satisfy. Define the necessary separation constants.

34. Consider a two-dimensional Cartesian coordinate system and a two-dimensional uv-system with the coordinates related by

$$x = -uv$$
$$y = (1/2)(u^2 - v^2).$$

In general, Laplace's equation in two dimensions can be written as

$$\nabla^2 \Psi(q_1, q_2) = \frac{1}{h_1 h_2} \left[\frac{\partial}{\partial q_1} \frac{h_2}{h_1} \frac{\partial}{\partial q_1} + \frac{\partial}{\partial q_2} \frac{h_1}{h_2} \frac{\partial}{\partial q_2} \right] \Psi(q_1, q_2) =,$$

with

$$h_i^2 = \left(\frac{\partial x}{\partial q_i} \right)^2 + \left(\frac{\partial y}{\partial q_i} \right)^2.$$

(a) In the xy-plane, sketch lines of constant u and constant v.

(b) Express Laplace's equation using the uv-coordinates.

(c) Use the method of separation of variables to separate Laplace's equation in the uv-system and obtain the general solution for $\Psi(u, v)$.

35. Let $n(y, t)$ represent the number density (the number of particles per unit volume) of neutrons in a slab of nuclear reactor material. Take the slab to be infinite in the x and z-directions, so that there is no variation in the number density with x or z. The motion of the neutrons is described by a diffusion coefficient D. The neutrons can also collide with atoms in the slab, creating unstable isotopes. When an unstable isotope decays, more neutrons are released than were needed to make the atom's nucleus unstable. This results in a source of neutrons that is accounted for by a collision frequency ν. Both D and ν are positive, real numbers. The diffusion and source effects can be combined to give a partial differential equation that describes the evolution of $n(y, t)$:

$$D \frac{\partial^2 n(y, t)}{\partial y^2} + \nu\, n(y, t) = \frac{\partial n(y, t)}{\partial t}.$$

At $y = 0$ and $y = L$, the neutrons can easily leave the slab, so the number density is always zero at these boundaries.

(a) Using the method of separation of variables, obtain a series solution for $n(y, t)$ that obeys the above boundary conditions and is valid for an arbitrary initial condition, i.e., $n(y, 0) = n_o(y)$.

(b) Using your answer to part (a), evaluate the series coefficients for the initial condition $n_o(y) = N_o \delta(y - L/2)$.

(c) The solution for an arbitrary initial condition can be formulated as a Green's function integral. What is this Green's function, $g(y|\xi, t)$? Express the solution $n(y, t)$ in terms of a Green's function integral and show that $n(y, t)$ goes to your answer to part (b) if $n_o(y) \rightarrow N_o \delta(y - L/2)$.

(d) Using your solution to part (a), determine the thickness L of the slab, which results in a nontrivial, steady-state solution for $n(y, t)$ as $t \rightarrow \infty$.

(e) What are the consequences of having a slab thickness greater than the value you calculated in part (d)?

12

INTEGRAL EQUATIONS

Differential equations have unknown functions or dependent variables inside differential operators. In contrast, integral equations have unknown functions or dependent variables inside integral operations. For example,

$$\frac{d^2}{dx^2}\phi(x) = \phi(x) \tag{12.1}$$

is a typical differential equation, with one dependent variable ϕ. This differential equation can be easily converted to an integral equation using a Green's function argument. The Green's function solution to Equation 12.1 takes the form

$$\phi(x) = \int_{-\infty}^{\infty} d\xi \, \phi(\xi)g(x|\xi), \tag{12.2}$$

where $g(x|\xi)$ satisfies

$$\frac{d^2}{dx^2}g(x|\xi) = \delta(x - \xi). \tag{12.3}$$

Equation 12.2 is an integral equation because the dependent variable ϕ appears inside the integral operation, and there are no differential operations.

In this chapter, we discuss four general classes of integral equations, the connection between differential and integral equations and several methods of solution. In Appendix G, a brief summary of the calculus of variations is given. This is a technique that uses the structure of an integral equation and minimizes the value of a definite integral with respect to an initially unknown function embedded inside the integral itself.

12.1   CLASSIFICATION OF LINEAR INTEGRAL EQUATIONS

The integral equations discussed in this chapter are all linear in the dependent variable. They fall into four categories: Fredholm equations of the first and second kind, and Volterra equations of the first and second kind. These classifications are important because the approaches to solving these equations are different in each case.

12.1.1   Fredholm Equations of the First Kind

In a Fredholm equation of the first kind, the dependent variable ϕ exists only inside the integral operation, and the limits of integration are constants. For example,

$$f(x) = \int_a^b dx' \, \phi(x')k(x, x') \tag{12.4}$$

is a Fredholm equation of the first kind. Both $f(x)$ and $k(x, x')$ are known functions of the independent variable x. The function $k(x, x')$ is referred to as a "kernel."

12.1.2   Fredholm Equations of the Second Kind

Fredholm equations of the second kind differ from Fredholm equations of the first kind in that the dependent variable appears both inside and outside the integral operation. For example,

$$\phi(x) = f(x) + \lambda \int_a^b dx' \, \phi(x')k(x, x') \tag{12.5}$$

is a Fredholm equation of the second kind. Again, $\phi(x)$ is the unknown function, but both $f(x)$ and the kernel $k(x, x')$ are known. The limits of the integral, a and b, and λ are all constants.

12.1.3   Volterra Equations of the First Kind

Volterra equations of the first kind are similar to Fredholm equations of the first kind, except the limits of integration are not constants. The unknown function still appears only inside the integral. Therefore,

$$f(x) = \int_a^x dx' \, \phi(x')k(x, x') \tag{12.6}$$

is an example of a Volterra equation of the first kind, because the upper limit of the integral is the variable x.

12.1.4   Volterra Equations of the Second Kind

Finally, a Volterra equation of the second kind has the unknown function both inside and outside the integral operation, and the limits of the integral are not constants. The

equation

$$\phi(x) = f(x) + \int_a^x dx'\, \phi(x')k(x, x') \tag{12.7}$$

is an example of a Volterra equation of the second kind.

12.2 THE CONNECTION BETWEEN DIFFERENTIAL AND INTEGRAL EQUATIONS

Differential equations can be converted to integral equations. In this process, the boundary conditions associated with the differential equation are automatically incorporated into the integral equation. Thus, unlike differential equations, integral equations will have unique solutions without specifying additional conditions.

Let's see how a ordinary second-order, linear differential equation can be converted to an integral equation. We start with the most general form

$$\frac{d^2 y(x)}{dx^2} + A(x)\frac{dy(x)}{dx} + B(x)y(x) = s(x), \tag{12.8}$$

where the functions $A(x)$, $B(x)$ and $s(x)$ are presumed to be known. Because this is a second-order equation, two boundary conditions are required to make the solution $y(x)$ unique. For this case, we will specify $y(x)$ and its derivative at the point $x = a$:

$$y(a) = y_o \tag{12.9}$$

and

$$\left.\frac{dy(x)}{dx}\right|_a = y_o'. \tag{12.10}$$

To convert Equation 12.8 into integral form, all the differential operations must be removed by successive integrations. The first such integration is accomplished by changing the variable in Equation 12.8 from x to x', and then operating on both sides with

$$\int_a^x dx'. \tag{12.11}$$

After this integration, Equation 12.8 becomes

$$\left.\frac{dy(x')}{dx'}\right|_x - \left.\frac{dy(x')}{dx'}\right|_a + \int_a^x dx'\, A(x')\frac{dy(x')}{dx'} \tag{12.12}$$

$$+ \int_a^x dx'\, B(x')y(x') = \int_a^x dx'\, s(x').$$

Applying the boundary condition of Equation 12.10 and integrating the first integral on the LHS by parts gives

$$\frac{dy(x)}{dx} - y_o' + A(x')y(x')\Big|_a^x - \int_a^x dx' \frac{dA(x')}{dx'} y(x')$$

$$+ \int_a^x dx' B(x')y(x') = \int_a^x dx' s(x'). \tag{12.13}$$

Using the other boundary condition given in Equation 12.9, this equation can be rewritten as

$$\frac{dy(x)}{dx} - y_o' + A(x)y(x) - A(a)y_o$$

$$+ \int_a^x dx' \left[B(x') - \frac{dA(x')}{dx'} \right] y(x') = \int_a^x dx' s(x'). \tag{12.14}$$

One more integration is necessary to remove the derivative in the first term. To do this, the variable x is changed to x'' and the equation is operated on by

$$\int_a^x dx''. \tag{12.15}$$

The dummy integration variable is labeled x'' to avoid confusion with x', the dummy variable of the first integration. This integration, and additional application of the boundary conditions, transforms Equation 12.13 to

$$y(x) - y_o - y_o'(x - a) + \int_a^x dx'' A(x'')y(x'') - A(a)y_o(x - a)$$

$$+ \int_a^x dx'' \int_a^{x''} dx' \left[B(x') - \frac{dA(x')}{dx'} \right] y(x') \tag{12.16}$$

$$= \int_a^x dx'' \int_a^{x''} dx' s(x').$$

Using the double integral identity

$$\int_a^x dx'' \int_a^{x''} dx' f(x') = \int_a^x dx' (x - x') f(x'), \tag{12.17}$$

which is proven in Appendix C, this equation becomes

$$y(x) - y_o - y_o'(x - a) + \int_a^x dx'' A(x'')y(x'') - A(a)y_o(x - a)$$

$$+ \int_a^x dx' (x - x') \left[B(x') - \frac{dA(x')}{dx'} \right] y(x') = \int_a^x dx'(x - x')s(x'). \tag{12.18}$$

This equation can be rearranged as

$$y(x) = y_o + y_o'(x - a) + A(a)y_o(x - a) + \int_a^x dx'(x - x')s(x')$$

$$+ \int_a^x dx' \left\{ (x - x') \left[\frac{dA(x')}{dx'} - B(x') \right] - A(x') \right\} y(x'), \quad (12.19)$$

which is in the form of a Volterra equation of the second kind, like Equation 12.7. The unknown function is

$$\phi(x) = y(x), \tag{12.20}$$

the known function is

$$f(x) = y_o + y_o'(x - a) + A(a)y_o(x - a) + \int_a^x dx'(x - x')s(x'), \tag{12.21}$$

and the kernel is

$$k(x, x') = (x - x') \left[\frac{dA(x')}{dx'} - B(x') \right] - A(x'). \tag{12.22}$$

The solution to Equation 12.19 is identical to the solution of Equation 12.8, when the boundary conditions given by Equations 12.9 and 12.10 are applied. To work with the differential equation, the boundary conditions must be specified in addition to the differential equation itself. With this integral equation, however, no additional information is necessary to arrive at the same unique solution. The boundary conditions are built in.

Example 12.1   As a specific example of the conversion of a differential equation to an integral equation, consider the differential equation for a harmonic oscillator,

$$\frac{d^2 y(x)}{dx^2} + k_o^2 y(x) = 0, \tag{12.23}$$

with the boundary conditions

$$y(0) = y(1) = 0. \tag{12.24}$$

Notice we cannot use the general result derived in the last section, because the form of the boundary conditions is different in this case. Here we have specified the unknown function's value at two end points. Before, we specified the function and its derivative at a single point.

As we did before, we will change $x \rightarrow x'$ and then apply the operator

$$\int_0^x dx' \tag{12.25}$$

to both sides of Equation 12.23 to give

$$\left.\frac{dy(x')}{dx'}\right|_{x'=x} - \left.\frac{dy(x')}{dx'}\right|_{x'=0} + k_o^2 \int_0^x dx' \, y(x') = 0. \tag{12.26}$$

The $(dy/dx)|_0$ term on the LHS is a constant, but it is not one of the boundary conditions. We label it as the constant y_0', but remember that it is not yet known. We integrate again, by transforming $x \rightarrow x''$, and applying the operator

$$\int_0^x dx'' \tag{12.27}$$

to both sides to obtain

$$y(x) - y(0) - y_0'x + k_o^2 \int_o^x dx'' \int_0^{x''} dx' \, y(x') = 0. \tag{12.28}$$

The quantity $y(0)$ is one of the boundary conditions. Setting $y(0) = 0$ and using the double integral identity of Equation 12.17 gives

$$y(x) - y_0'x + k_o^2 \int_0^x dx' \, (x - x') y(x') = 0. \tag{12.29}$$

We are not finished yet, because the constant y_0' is not known. Also, we have not made use of the second boundary condition at $x = 1$. These two things are taken care of at once by evaluating Equation 12.29 at $x = 1$:

$$y(1) - y_o' + k_o^2 \int_0^1 dx' \, (1 - x') y(x') = 0. \tag{12.30}$$

Because $y(1) = 0$, this equation allows y_o' to be evaluated as

$$y_0' = k_o^2 \int_0^1 dx' \, (1 - x') y(x'). \tag{12.31}$$

Equation 12.29 becomes

$$y(x) = k_o^2 x \int_0^1 dx' \, (1 - x') y(x') - k_o^2 \int_0^x dx' \, (x - x') y(x'). \tag{12.32}$$

This is clearly an integral equation. All the differential operators have been removed, and both the boundary conditions have been incorporated. The unknown function appears both inside and outside the integral operations, but one integral has constant

limits and the other has the independent variable as an upper limit. This appears to be a mixture of a Fredholm equation of the second kind and a Volterra equation of the second kind. A bit more work clears up this confusion.

The integration from $x' = 0$ to $x' = 1$ can be broken up into two parts. Because the independent variable x ranges from 0 to 1, Equation 12.32 can be rewritten as

$$y(x) = k_o^2 \int_0^x dx' \, x(1 - x') y(x') + k_o^2 \int_x^1 dx' \, x(1 - x') y(x')$$
$$- k_o^2 \int_0^x dx' \, (x - x') y(x'). \tag{12.33}$$

The two integrals from $x' = 0$ to $x' = x$ can be combined:

$$y(x) = k_o^2 \int_0^x dx' x'(1 - x) y(x') + k_o^2 \int_x^1 dx' x(1 - x') y(x'). \tag{12.34}$$

Equation 12.34 is now in the form of a Fredholm equation of the second kind:

$$y(x) = k_o^2 \int_0^1 dx' k(x, x') y(x'), \tag{12.35}$$

with

$$k(x, x') = \begin{cases} x'(1 - x) & 0 \le x' \le x \\ x(1 - x') & x \le x' \le 1 \end{cases}. \tag{12.36}$$

Notice that the independent variable x is a constant inside the integral operation, while the x' is the variable of integration.

The kernel for this problem is plotted in Figure 12.1. This kernel should look familiar to you. The differential equation for the string problem of the last chapter was solved using a Green's function, which was essentially identical to the kernel described by Equation 12.36. The close link between this integral equation and the Green's function solution isn't too surprising when you consider Equation 12.23 from a Green's function point of view. Rewrite Equation 12.23 as

$$\frac{d^2 y(x)}{dx^2} = -k_o^2 y(x). \tag{12.37}$$

K(x,x′)
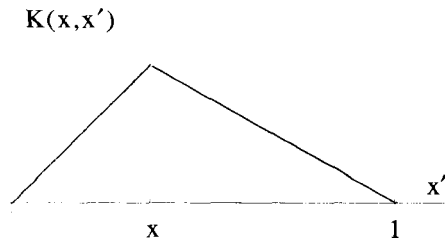
Figure 12.1 Kernel for the Harmonic Problem

If the RHS of this equation is looked at as the drive, then the Green's function solution for $y(x)$ can be written as

$$y(x) = -k_o^2 \int_0^1 d\xi \, g(x|\xi) y(\xi), \qquad (12.38)$$

where $g(x|\xi)$ is the Green's function associated with the $d^2 y(x)/dx^2$ differential operator. That is, $g(x|\xi)$ is the solution of

$$\frac{d^2 g(x|\xi)}{dx^2} = \delta(x - \xi), \qquad (12.39)$$

with the same homogeneous boundary conditions as $y(x)$

$$g(0|\xi) = g(1|\xi) = 0. \qquad (12.40)$$

This Green's function solution, Equation 12.38, is in the form of an integral equation for $y(x)$. It is identical to the integral equation we obtained by converting the original differential equation, Equation 12.37, to an integral equation with

$$g(x|x') = -k(x, x'). \qquad (12.41)$$

12.3   METHODS OF SOLUTION

In this section, several methods of solution for integral equations are presented. The different forms of integral equations generally require different solution techniques. For example, one takes a different approach for Volterra equations of the first kind than for Fredholm equations of the second kind.

In these discussions, it is useful to define two classifications of kernels, which we dub "translationally invariant" and "causal" kernels. A translationally invariant kernel has the form

$$k(x, x') = k(x - x'). \qquad (12.42)$$

The motivation for this terminology is fairly evident, when you consider the connection with the Green's function kernels we discussed in the previous section. For a translationally invariant kernel, if the drive is displaced by a certain amount, the response will not change its character, but will just be displaced by the same amount. A causal kernel obeys the condition

$$k(x, x') = 0 \qquad x < x', \qquad (12.43)$$

which again using the Green's function analogy, means that a response does not occur before the drive. Several of our solution techniques require kernels which obey one or both of these conditions.

12.3.1   Fourier Transform Solutions

Recall, the standard Fourier transform operation is defined by the pair of expressions

$$\underline{F}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt\, e^{-i\omega t} f(t) \tag{12.44}$$

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega\, e^{i\omega t} \underline{F}(\omega). \tag{12.45}$$

A Fourier transform approach works well for finding solutions to Fredholm equations of the first kind, which have both translationally invariant kernels and infinite limits of integration:

$$f(t) = \int_{-\infty}^{\infty} d\tau\, k(t - \tau) y(\tau). \tag{12.46}$$

In this expression, $f(t)$ and $k(t - \tau)$ are known functions, and $y(t)$ is unknown. Assume that the Fourier transform of each of these functions exists:

$$f(t) \leftrightarrow \underline{F}(\omega) \tag{12.47}$$

$$k(t) \leftrightarrow \underline{K}(\omega) \tag{12.48}$$

$$y(t) \leftrightarrow \underline{Y}(\omega). \tag{12.49}$$

Notice, because the kernel is translationally invariant, we can describe its argument using only a single variable.

Because Equation 12.46 is a convolution of $k(t)$ and $y(t)$, its Fourier transform is easy to evaluate:

$$\underline{F}(\omega) = \sqrt{2\pi}\, \underline{K}(\omega)\underline{Y}(\omega). \tag{12.50}$$

Solving for $\underline{Y}(\omega)$ gives

$$\underline{Y}(\omega) = \frac{\underline{F}(\omega)}{\sqrt{2\pi}\,\underline{K}(\omega)}. \tag{12.51}$$

The solution of $y(t)$ can now be obtained with a Fourier inversion:

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, e^{i\omega t} \frac{\underline{F}(\omega)}{\underline{K}(\omega)}. \tag{12.52}$$

12.3.2   Laplace Transform Solutions

The Laplace transform approach works for solving Volterra equations of the first kind with causal, translationally invariant kernels and integration limits that range from

$\tau = 0$ to $\tau = t$:

$$f(t) = \int_0^t d\tau\, k(t - \tau) y(\tau). \tag{12.53}$$

Again, $f(t)$ and the kernel are known functions, while $y(t)$ is the unknown. We assume $f(t)$, $y(t)$, and $k(t)$ are all zero for $t < 0$ and possess the Laplace transforms

$$f(t) \leftrightarrow \underline{F}(\underline{s}) \tag{12.54}$$

$$k(t) \leftrightarrow \underline{K}(\underline{s}) \tag{12.55}$$

$$y(t) \leftrightarrow \underline{Y}(\underline{s}). \tag{12.56}$$

Recall that the standard Laplace operation is defined by the equation pair

$$\underline{F}(\underline{s}) = \int_0^\infty dt\, e^{-st} f(t) \tag{12.57}$$

$$f(t) = \frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s}\, e^{st} \underline{F}(\underline{s}), \tag{12.58}$$

with the Laplace contour $\mathcal{L}$ to the right of all the poles of $\underline{F}(\underline{s})$.

As it stands now, the integral in Equation 12.53 is not a convolution integral because the upper limit of the integration is t rather than infinity. This can easily be corrected, since we assumed a causal kernel. Because $k(t - \tau) = 0$ for $t < \tau$, we can extend the upper limit of the integral to $+\infty$ without changing the value of the integral. Because $y(\tau)$ is zero for $\tau < 0$ the lower limit can be extended to $-\infty$ without changing the value:

$$f(t) = \int_0^t d\tau\, k(t - \tau) y(\tau) = \int_{-\infty}^\infty d\tau\, k(t - \tau) y(\tau). \tag{12.59}$$

The integral equation is now in the form of a standard convolution and applying the Laplace transform operation to both sides gives

$$\underline{F}(\underline{s}) = \underline{K}(\underline{s}) \underline{Y}(\underline{s}). \tag{12.60}$$

Solving for $Y(\underline{s})$ gives

$$\underline{Y}(\underline{s}) = \frac{\underline{F}(\underline{s})}{\underline{K}(\underline{s})}. \tag{12.61}$$

The solution for $y(t)$ is obtained with a Laplace inversion:

$$y(t) = \frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s}\, e^{st} \frac{\underline{F}(\underline{s})}{\underline{K}(\underline{s})}. \tag{12.62}$$

In this expression, the Laplace contour must be to the right of all the poles of the quantity $[\underline{F}(\underline{s})/\underline{K}(\underline{s})]$.

12.3.3 Series Solutions

Fredholm equations of the second kind can be solved by developing what is known as a Neumann series solution. In quantum mechanics, this process is called perturbation theory. The general form of a Fredholm equation of the second kind is

$$y(x) = f(x) + \lambda \int_a^b dx' k(x, x')y(x'), \tag{12.63}$$

where $f(x)$ and $k(x, x')$ are known functions, λ is a constant, and $y(x)$ is the unknown.

The constant λ is assumed to be "small" in some sense. To start, we will assume that λ is small enough that

$$|f(x)| \gg \left| \lambda \int_a^b dx' k(x, x')y(x') \right| \tag{12.64}$$

for all $a < x < b$. This really does not pin down the magnitude of λ very well, because the integral on the RHS of Equation 12.64 cannot be evaluated without knowing $y(x)$ in advance. Notice how $f(x)$ cannot be zero anywhere. If it were, Equation 12.64 could never be satisfied, regardless of the magnitude of λ. Colloquially, we describe this requirement by saying the method requires a nonzero "seed" function.

If Equation 12.64 is satisfied, an iterative approach can be employed to obtain a solution for $y(x)$ to arbitrary orders of λ. As a first approximation, the solution can be given by

$$y_0(x) = f(x). \tag{12.65}$$

We call this the zeroth-order solution because λ does not appear in this expression. An improvement on this approximate solution is made if we substitute the zeroth-order solution into the integral of Equation 12.63 and then solve for a new value of $y(x)$:

$$\begin{aligned} y_1(x) &= f(x) + \lambda \int_a^b dx' k(x, x')y_o(x') \\ &= f(x) + \lambda \int_a^b dx' k(x, x')f(x'). \end{aligned} \tag{12.66}$$

We call $y_1(x)$ the first-order solution, because it has a term with λ raised to the first power. Now you probably see what we are doing. To get the second-order solution, we substitute $y_1(x)$ back into Equation 12.63 and solve for a new $y(x)$:

$$\begin{aligned} y_2(x) &= f(x) + \lambda \int_a^b dx' k(x, x')y_1(x') \\ &= f(x) + \lambda \int_a^b dx' k(x, x')f(x') \\ &\quad + \lambda^2 \int_a^b dx' k(x, x') \int_a^b dx'' k(x', x'')f(x''). \end{aligned} \tag{12.67}$$

And of course, you guessed it, this is called a second-order solution because the highest power of λ that appears is λ^2. This process can be carried to any order. The nth-order solution is

$$y_n(x) = f(x) + \lambda \int_a^b dx' k(x, x') y_{n-1}(x')$$

$$= f(x) + \lambda \int_a^b dx' k(x, x') f(x') \tag{12.68}$$

$$+ \lambda^2 \int_a^b dx' \int_a^b dx'' \, k(x, x') k(x', x'') f(x'')$$

$$+ \lambda^3 \int_a^b dx' \int_a^b dx'' \int_a^b dx''' k(x, x') k(x', x'') k(x'', x''') f(x''')$$

$$\vdots$$

$$+ \lambda^n \int_a^b dx' \cdots \int_a^b dx^{n'} k(x, x') \cdots k(x^{(n-1)'}, x^{n'}) f(x^{n'}).$$

In this expression $x^{n'}$ is used to indicate an x with n primes.

Now our hope is that

$$y(x) = \lim_{n \to \infty} y_n(x). \tag{12.69}$$

That is, if we keep performing the iterative process of substituting our approximate result for $y(x)$ back into the equation, we will eventually converge on the correct solution. The standard convergence tests should be applied to check that the resulting infinite sum converges. Obviously, λ must be small enough for this to occur.

Example 12.2 Most engineering and physics problems formulate naturally as differential equations as opposed to integral equations. That is why more emphasis, in this book and in others, is placed on solutions to differential equations rather than integral equations. There are some problems, however, that do formulate naturally as integral equations. The quantum mechanical scattering problem is one.

Consider an electron incident on a scattering center, such as the hydrogen nucleus sketched in Figure 12.2. The scattering center is modeled by an electrostatic potential $V(\bar{r})$ that is a scalar function of position. The electron is modeled by a wave function $\psi(\bar{r})$, and its interaction with the potential is described by the Schrödinger wave

Figure 12.2 The Quantum Mechanical Scattering Problem

equation:

$$\left[-\frac{\hbar^2 \nabla^2}{2m} + V(\bar{r}) \right] \psi(\bar{r}) = E\psi(\bar{r}). \tag{12.70}$$

Here E is the energy of the electron and m is its mass. If we define the constant $k^2 = 2mE/\hbar^2$, Equation 12.70 can be rewritten as

$$\left(\nabla^2 + k^2 \right) \psi(\bar{r}) = \frac{2m}{\hbar^2} V(\bar{r})\psi(\bar{r}). \tag{12.71}$$

This last equation is a linear differential equation with $\psi(\bar{r})$ as the dependent variable. A Green's function approach to solving this equation gives the result

$$\psi(\bar{r}) = e^{i\bar{k}\cdot\bar{r}} + \frac{2m}{\hbar^2} \int d^3\bar{r}' \, g(\bar{r}|\bar{r}')V(\bar{r}')\psi(\bar{r}'). \tag{12.72}$$

The first term on the RHS is the solution for $\psi(\bar{r})$ when $V(\bar{r})$ is zero, the homogeneous version of Equation 12.71. The second term is an integral over all space involving the Green's function $g(\bar{r}|\bar{r}')$. The Green's function itself is a solution to

$$\left(\nabla^2 + k^2 \right) g(\bar{r}|\bar{r}') = \delta^3(\bar{r} - \bar{r}'). \tag{12.73}$$

It will be assumed that the Green's function vanishes at infinity. Applying this boundary condition gives

$$g(\bar{r}|\bar{r}') = \frac{e^{ik|\bar{r}-\bar{r}'|}}{4\pi|\bar{r} - \bar{r}'|}. \tag{12.74}$$

If $\psi(\bar{r})$ were not sitting inside the Green's function integral of Equation 12.72, we would already have our desired solution. But since it does appear both inside and outside of the integral, we resort to the Neumann series method described earlier. The potential $V(\bar{r})$ is assumed to be a small perturbation to the problem. The zero-order solution is just the wave function for a free electron

$$\psi_0(\bar{r}) = e^{i\bar{k}\cdot\bar{r}}. \tag{12.75}$$

We obtain the first-order solution by forcing this solution back into Equation 12.72:

$$\psi_1(\bar{r}) = e^{i\bar{k}\cdot\bar{r}} + \frac{2m}{\hbar^2} \int d^3\bar{r}' g(\bar{r}|\bar{r}')V(\bar{r}')e^{i\bar{k}\cdot\bar{r}'}. \tag{12.76}$$

This solution is referred to as the first Born approximation. Higher-order approximations follow the Neumann series solution method.

You may have noticed something a little odd about this example. In the derivation of the Neumann series, the constant λ needed to be small enough for the series to converge. In this example, however, the constant λ is nowhere to be seen. The convergence of this solution is controlled by the magnitude of $V(\bar{r})$.

12.3.4 Separable Kernel Solutions

Fredholm equations of the second kind can be solved by the method of separable kernels, if the kernel can be made equal to the sum of several products of two functions. In mathematical terms, the kernel is separable when

$$k(x, x') = \sum_{j=1,2,\cdots}^{N} M_j(x)N_j(x'). \tag{12.77}$$

Notice $M_j(x)$ and $N_j(x')$ are functions of only one variable. This separation can be accomplished for a surprisingly large number of kernels, even, as we shall see, a translationally invariant one.

If the kernel separates, we can write a Fredholm equation of the second kind as

$$y(x) = f(x) + \lambda \int_a^b dx' \sum_{j=1}^{N} M_j(x)N_j(x')y(x'). \tag{12.78}$$

In this equation, $f(x)$ and the $M_j(x)$ and $N_j(x)$ functions are known, and λ is a constant. Interchanging the order of integration and summation gives

$$y(x) = f(x) + \lambda \sum_{j=1}^{N} M_j(x) \int_a^b dx' N_j(x')y(x'). \tag{12.79}$$

Now this integration has no x-dependence. If we define

$$C_j = \int_a^b dx' N_j(x')y(x'), \tag{12.80}$$

Equation 12.78 can be rewritten as

$$y(x) = f(x) + \lambda \sum_{j=1}^{N} M_j(x)C_j. \tag{12.81}$$

Keep in mind, the constants C_j are still unknowns because $y(x)$ is unknown.

Equation 12.81 is in a mixed form. The unknown elements are both the explicit $y(x)$ and the C_j's. The known elements are $f(x)$ and the $M_j(x)$ functions. Some terms have x-dependence and some terms are constants. We can generate a more consistent form if we operate on both sides of Equation 12.81 by

$$\int_a^b dx\, N_i(x), \tag{12.82}$$

which yields

$$\int_a^b dx\, N_i(x)y(x) = \int_a^b dx\, N_i(x)f(x) + \lambda \sum_{j=1}^{N} C_j \int_a^b dx\, N_i(x)M_j(x). \qquad (12.83)$$

According to Equation 12.80, the integral on the LHS is just the unknown constant C_i. The first integral on the RHS can be defined as a new, known constant with an i subscript:

$$B_i = \int_a^b dx\, N_i(x)f(x). \qquad (12.84)$$

The last integral generates the known elements of a matrix $[A]$:

$$A_{ij} = \int_a^b dx\, N_i(x)M_j(x). \qquad (12.85)$$

Equation 12.83 can now be written in terms of known and unknown constants as

$$C_i = B_i + \lambda \sum_{j=1}^{N} C_j A_{ij} \qquad (12.86)$$

or in matrix notation as

$$[C] = [B] + \lambda[A][C], \qquad (12.87)$$

where the matrices $[B]$ and $[C]$ can be viewed as column matrices made up of the elements of B_i and C_i.

Working in matrix notation, and rearranging Equation 12.87 gives

$$[B] = \{[1] - \lambda[A]\}[C], \qquad (12.88)$$

where $[1]$ is the $N \times N$ diagonal, identity matrix. Solving for $[C]$ gives

$$[C] = \{[1] - \lambda[A]\}^{-1}[B]. \qquad (12.89)$$

To determine the elements of $[C]$ requires an inversion of the matrix

$$[1] - \lambda[A]. \qquad (12.90)$$

Once these elements of $[C]$ are determined, $y(x)$ is generated quite simply using Equation 12.81:

$$y(x) = f(x) + \lambda \sum_{j=1}^{N} C_j M_j(x). \qquad (12.91)$$

This matrix notation points out a very interesting result for these types of problems. If $f(x) = 0$, then all the elements of the $[B]$ matrix are also zero. Therefore, from Equation 12.88, it can be seen that a nonzero $[C]$ matrix, and therefore a nonzero $y(x)$, can exist only if

$$|[1] - \lambda [A]|_{det} = 0. \tag{12.92}$$

Given the $[A]$ matrix, Equation 12.92 will be satisfied only for specific values of λ. A nonzero solution to the original integral equation will exist only for these λ_n eigenvalues, and the solutions $y_n(x)$ are eigenfunctions of the integral equation:

$$\int_a^b dx' \, k(x, x') \, y_n(x') = \frac{1}{\lambda_n} y_n(x). \tag{12.93}$$

Example 12.3   It is possible to separate kernels that do not appear, at first glance, to be separable. As an example, consider the translationally invariant kernel

$$k(t, t') = \cos(t - t'). \tag{12.94}$$

This kernel can be separated with the trigonometric identity

$$\cos(t - t') = \cos(t)\cos(t') + \sin(t)\sin(t'), \tag{12.95}$$

so that

$$\cos(t - t') = \sum_{j=1}^{2} M_j(t) N_j(t'), \tag{12.96}$$

where

$$
\begin{aligned}
M_1(t) &= \cos(t) & M_2(t) &= \sin(t) \\
N_1(t') &= \cos(t') & N_2(t') &= \sin(t')
\end{aligned}
\tag{12.97}
$$

EXERCISES FOR CHAPTER 12

1. Derive the Fredholm integral equation that corresponds to

$$\frac{d^2 y(x)}{dx^2} - y(x) = 0,$$

   subject to the boundary conditions $y(1) = y(-1) = 0$. How does the integral equation change if the boundary conditions are changed to $y(1) = dy/dx|_1 = 1$?

2. Convert the following integral equation to a differential equation:

$$y(x) = \int_0^x dx'(x^2 - xx')y(x').$$

Identify all the boundary conditions included in the integral equation, which must be specified to make the solution to the differential equation unique.

3. Given the Fredholm equation of the second kind,

$$y(x) = 1 - \int_{-1}^{1} dt\, K(x, t) y(t),$$

where

$$k(x, t) = \begin{cases} (1 - x)(t + 1)/2 & -1 < t < x \\ (1 - t)(x + 1)/2 & x < t < 1 \end{cases},$$

obtain the associated second-order differential equation and the boundary conditions.

4. Convert the integral equation

$$y(x) = 1 - x + \int_{0}^{x} dt\, (x - t) y(t)$$

to a differential equation, with the appropriate boundary conditions.

5. Convert the integral equation

$$\phi(x) = 1 + \lambda^2 \int_{0}^{x} dt\, (x - t) \phi(t)$$

to a differential equation and determine the appropriate boundary conditions.

6. Use Fourier techniques to solve the following Fredholm integral equation for $\phi(x)$:

$$e^{-x^2} = \int_{-\infty}^{\infty} dt\, \phi(t) e^{-(x-t)^2}.$$

7. Use Laplace techniques to solve the following integral equation for $\phi(x)$:

$$\phi(x) = x - \int_{0}^{x} dt\, (t - x) \phi(x).$$

8. Using the Neumann series approach, solve

$$\phi(x) = 1 - 2 \int_{0}^{x} dt\, t \phi(t)$$

to obtain $\phi(x) = e^{-x^2}$.

9. Consider the Fredholm equation

$$\phi(x) = \lambda \int_{-\pi}^{\pi} dy\,[1 + \sin(x + y)]\phi(y).$$

   (a) Using the method of separable kernels, determine the values for λ which allow solution for $\phi(x)$.
   (b) Determine all the nonzero solutions for $\phi(x)$.

10. Consider the differential equation

$$\frac{dy(x)}{dx} + \lambda_o^2 x^2 y(x) = 0,$$

    with the boundary condition $y(0) = 1$.
    (a) Convert this differential equation and its boundary condition to an integral equation.
    (b) Using the Neumann series approach, obtain a solution to the integral equation assuming small λ_o. Can you place this solution in closed form, i.e., not in terms of an infinite series?

11. Solve

$$\phi(x) = x + \int_0^1 dt\,(1 + xt)\phi(t)$$

    by three methods,
    (a) the Neumann series technique.
    (b) the separable kernel technique.
    (c) educated guessing.

12. Find the possible eigenvalues and the eigenfunction solutions of the following integral equations:

    i. $\phi(x) = \lambda \int_{-1}^{1} dt\,(x - t)^2 \phi(x).$

    ii. $\phi(x) = \lambda \int_{-1}^{1} dt\,(t - x)\phi(t)$

    iii. $\phi(x) = \lambda \int_{-1}^{1} dt\,(xt + x)\phi(t).$

    iv. $\phi(x) = \lambda \int_{0}^{1} dt\,(1 + xt)\phi(t).$

13

ADVANCED TOPICS IN COMPLEX ANALYSIS

This chapter continues our discussion of complex variables with the investigation of two important, advanced topics. First, the complications associated with multi-valued complex functions are explored. We will introduce the concepts of branch points, branch cuts, and Riemann sheets, which will allow us to continue to use the residue theorem and other Cauchy integral techniques on these functions. The second topic describes the method of steepest descent, a technique useful for generating approximations for certain types of contour integrals.

13.1 MULTIVALUED FUNCTIONS

Up to this point, we have concentrated on complex functions that are single valued. A function $\underline{w}(\underline{z})$ is a single-valued function of $\underline{z}$ if it generates one value of $\underline{w}$ for a given value of $\underline{z}$. The function

$$\underline{w} = \underline{z}^2 \tag{13.1}$$

is example of a single-valued function of $\underline{z}$. In contrast, the function

$$\underline{w} = \underline{z}^{1/2} \tag{13.2}$$

is an example of a multivalued function of $\underline{z}$. This is easy to see, because $\underline{z} = 1$ generates $\underline{w} = +1$ or $\underline{w} = -1$. Another point, $\underline{z} = 2i$, gives either $\underline{w} = \sqrt{2}\,e^{i\pi/4}$ or $\underline{w} = \sqrt{2}\,e^{i5\pi/4}$.

In this chapter, we will use the convention that the $\sqrt{}$ operation is performed only on a positive real quantity to produce a single, positive, real result. Operations like $(\ )^{1/2}$ can have positive, negative, or complex arguments and produce multivalued results.

509

13.1.1   Contour Integration of Multivalued Functions

In Chapter 6, special techniques were derived for performing integrals in the complex plane. Can we still use the residue theorem and the other Cauchy integral techniques on these multivalued functions? The answer is yes, if we are careful!

From our previous work, we know that if the complex function $\underline{w}(\underline{z})$ is analytic inside a closed contour C, we can write

$$\oint_C d\underline{z}\,\underline{w}(\underline{z}) = 0. \tag{13.3}$$

The key requirement here is that the contour must be closed. For this path to be closed, the variable of integration $\underline{z}$ has to return to its starting value. For a single-valued function, this obviously also brings $\underline{w}(\underline{z})$ back to its starting value. The same is not always true with a multivalued function, and this is the source of some difficulties, as you will see in the two examples that follow.

Example 13.1   The single-valued function

$$\underline{w} = \underline{z}^2 \tag{13.4}$$

is analytic everywhere in the complex $\underline{z}$ plane. Remember, for a function to be analytic, we must be able to write $\underline{w} = u(x, y) + iv(x, y)$, with $u(x, y)$ and $v(x, y)$ being continuous, real functions that obey the Cauchy-Riemann conditions. In this case,

$$u(x, y) = x^2 - y^2$$
$$v(x, y) = 2xy \tag{13.5}$$

and

$$\frac{\partial u}{\partial x} = 2x = \frac{\partial v}{\partial y}$$
$$\frac{\partial v}{\partial x} = 2y = -\frac{\partial u}{\partial y}. \tag{13.6}$$

So the Cauchy-Riemann conditions are satisfied for all $\underline{z}$. If the closed contour C is the unit circle centered at $\underline{z} = 0$, as shown in Figure 13.1, then according to the Cauchy integral theorem,

$$\oint_C d\underline{z}\,\underline{z}^2 = 0. \tag{13.7}$$

Let's be skeptical about this last result and explicitly perform the integration of Equation 13.7. On the contour C, $\underline{z} = e^{i\theta}$ and $d\underline{z} = ie^{i\theta}d\theta$. To go around the circle once and return $\underline{z}$ to its initial value, we let the range of θ vary from 0 to 2π. We could have made the limits on θ different, perhaps π to 3π, but the starting and ending
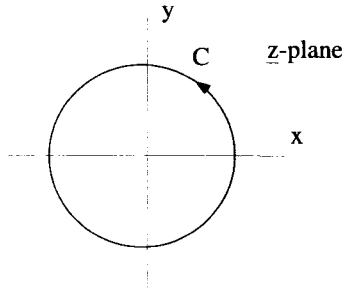
Figure 13.1   Contour for Unit Circle Integration

values of θ must differ by 2π so the circle is traversed exactly once. The integral becomes

$$\oint_C dz\,z^2 = \int_0^{2\pi} d\theta\, ie^{i\theta}e^{i2\theta}$$

$$= (1/3)e^{i3\theta}\Big|_0^{2\pi}$$

$$= (1/3)(e^{i6\pi} - 1)$$

$$= 0, \tag{13.8}$$

which is exactly the expected result. The integral around a closed path of this analytic function is zero. There are no surprises here.

Example 13.2   Now consider the multivalued function

$$w = z^{3/2}. \tag{13.9}$$

The single point $z = 1$ will give two different values for w: $w = 1$ and $w = -1$. A simple way to see this is to write z using an exponential representation

$$z = 1 = e^{i0} = e^{i2\pi} = e^{i4\pi} \cdots. \tag{13.10}$$

Notice that all these expressions identify the point $z = 1$, but do so using different polar angles. Now use the fact that $(e^{i\theta})^{3/2} = e^{3\theta/2}$ to obtain the corresponding values of w:

$$w = \begin{cases} e^{i0} = e^{i6\pi} = e^{i12\pi} \cdots = 1 \\ \qquad\qquad \text{or} \\ e^{i3\pi} = e^{i9\pi} = e^{i15\pi} \cdots = -1 \end{cases}. \tag{13.11}$$

Two different values for w are generated by this process.

Is this function analytic? If we write $\underline{z} = x + iy$, we have

$$\underline{w} = (x + iy)^{3/2}, \tag{13.12}$$

and with $\underline{w} = u(x, y) + iv(x, y)$,

$$u(x, y) = \frac{1}{2}(\underline{w} + \underline{w}^*) = \frac{1}{2}\left[(x + iy)^{3/2} + (x - iy)^{3/2}\right]$$

$$v(x, y) = \frac{1}{2i}(\underline{w} - \underline{w}^*) = \frac{1}{2i}\left[(x + iy)^{3/2} - (x - iy)^{3/2}\right]. \tag{13.13}$$

The partial derivatives become

$$\frac{\partial u}{\partial x} = \frac{3}{4}\left[(x + iy)^{1/2} + (x - iy)^{1/2}\right]$$

$$\frac{\partial u}{\partial y} = \frac{3i}{4}\left[(x + iy)^{1/2} - (x - iy)^{1/2}\right]$$

$$\frac{\partial v}{\partial x} = \frac{3}{4i}\left[(x + iy)^{1/2} - (x - iy)^{1/2}\right] \tag{13.14}$$

$$\frac{\partial v}{\partial y} = \frac{3}{4}\left[(x + iy)^{1/2} + (x - iy)^{1/2}\right],$$

and so the Cauchy-Riemann conditions are satisfied for all $\underline{z}$. This means the function given in Equation 13.9 is analytic everywhere in the $\underline{z}$-plane, and consequently its closed line integral around the unit circle shown in Figure 13.1 should be zero:

$$\oint_C d\underline{z}\,\underline{z}^{3/2} = 0. \tag{13.15}$$

Again, let's be skeptical about this result and check it explicitly by performing the same type of integration worked out in the previous example. As before, we set $\underline{z} = e^{i\theta}$, $d\underline{z} = ie^{i\theta}d\theta$, and let θ range from zero to 2π:

$$\int_0^{2\pi} d\theta\, ie^{i\theta}(e^{i\theta})^{3/2} = \int_0^{2\pi} d\theta\, ie^{i5\theta/2}$$

$$= \frac{2}{5}e^{i5\theta/2}\Big|_0^{2\pi}$$

$$= \frac{2}{5}[e^{i5\pi} - 1] \tag{13.16}$$

$$= -\frac{4}{5}.$$

Uh-oh! The Cauchy integral theorem appears to be violated. Our function is analytic inside C, but the integral is nonzero. Because the Cauchy integral theorem is the basis for complex integration using the residue theorem, its violation would be a big problem. Fortunately, there is a way to preserve the validity of the theorem.

Notice, although the 0 to 2π range of θ returns $\underline{z}$ to its starting value in the above integration, it does not return $\underline{w}$ to its initial value. When $\theta = 0$, $z = e^{i0} = 1$ and $\underline{w} = 1$, but when $\theta = 2\pi$, $z = e^{i2\pi} = 1$ and $\underline{w} = e^{i3\pi} = -1$. But if we go around the unit circle *twice*, that is $0 < \theta < 4\pi$, both $\underline{z}$ and $\underline{w}(\underline{z}) = \underline{z}^{3/2}$ will return to their starting values, and we obtain

$$
\begin{aligned}
\int_0^{4\pi} d\theta \, ie^{i\theta}(e^{i\theta})^{3/2} &= \int_0^{4\pi} d\theta \, ie^{i5\theta/2} \\
&= \frac{2}{5}e^{i5\theta/2}\bigg|_0^{4\pi} \\
&= \frac{2}{5}[e^{i10\pi} - 1] \\
&= 0.
\end{aligned}
\tag{13.17}
$$

It appears that the definition of closed integral contour must take on new meaning when we are using multivalued functions. We cannot say a contour is closed unless both the variable of integration and the function being integrated simultaneously return to their initial values. We will investigate this idea in detail in the sections that follow.

13.1.2   Riemann Sheets, Branch Points, and Branch Cuts

Branch points, branch cuts, and Riemann sheets are all concepts developed so that the Cauchy integral theorem, the residue theorem, and all the other tools developed for analytic functions will still work with multivalued, complex functions.

Riemann Sheets   The nature of multivalued functions can best be explained by looking at their mapping properties. Consider the simple function,

$$
\underline{w}(\underline{z}) = \underline{z}^{1/2}.
\tag{13.18}
$$

It is easy to see this function is multivalued. The single point $\underline{z} = e^{i\pi/4} = e^{i9\pi/4}$ maps onto two different points in the $\underline{w}$-plane, $\underline{w} = e^{i\pi/8}$ and $\underline{w} = e^{i9\pi/8}$. This mapping is shown in Figure 13.2.
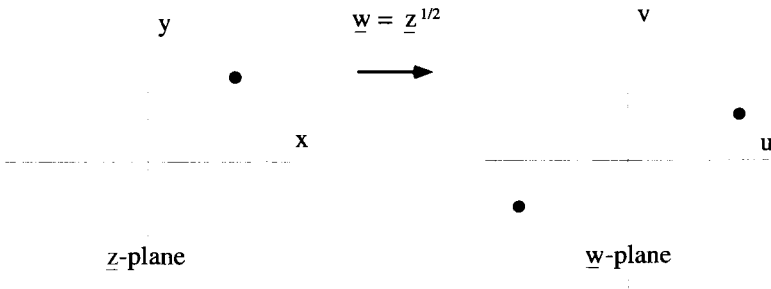
Figure 13.2   Mapping of a Multivalued Function
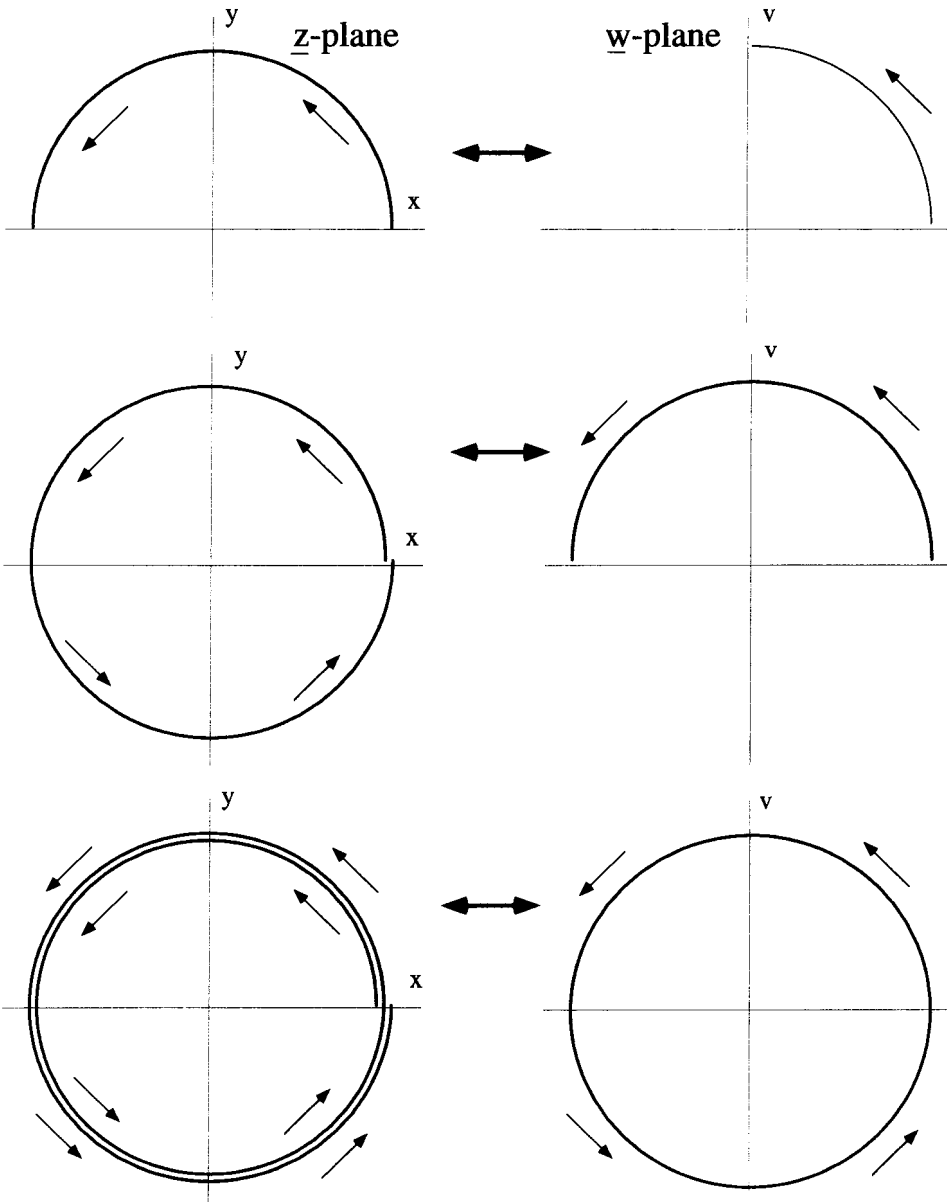
Figure 13.3   Mapping of $\underline{w} = \underline{z}^{1/2}$

Let's look at the mapping in more detail by determining how each of the points on a unit circle, centered around the origin of the $\underline{z}$-plane, gets mapped onto the $\underline{w}$-plane. On the circle, $\underline{z} = e^{i\theta}$ and $\underline{w} = e^{i\theta/2} \equiv e^{i\phi}$. This mapping is shown, divided into three steps, in Figure 13.3. At the start, we set $\theta = 0$, so $\underline{z} = 1$ and $\underline{w} = 1$. In the first

pair of pictures, we see that as the polar angle θ increases from 0 to π, ϕ increases from 0 to $\pi/2$. So $\underline{w}$ moves along its own unit circle, lagging $\underline{z}$ in angle by a factor of two. In the middle pair of pictures, θ continues to increase from π to 2π, so that $\underline{z}$ returns to $\underline{z} = e^{i2\pi} = 1$, its initial value. However, $\underline{w}$ ends up at $\underline{w} = e^{i\pi} = -1$, because ϕ has only increased to π. In the last pair of pictures, θ continues to increase from 2π to 4π, repeating the loop it has already made, and once again returns to its initial value of $\underline{z} = e^{i4\pi} = 1$. At the same time, $\underline{w}$ finally makes a complete circle and returns to its initial value, $\underline{w} = e^{i2\pi} = 1$. With the interpretation that a closed contour in the $\underline{z}$-plane must return both $\underline{z}$ and the function $\underline{w}(\underline{z})$ to their starting values, the unit circle in the $\underline{z}$-plane must be traversed twice in order to form a closed path for the function $\underline{w} = \underline{z}^{1/2}$.

To facilitate contour integration and other tasks that involve multivalued functions, it is useful to extend our concept of the complex plane to include what are commonly called Riemann sheets. Simply put, this is a way to force the mapping of a multivalued function to be one-to-one, so we can keep track of what contours are truly closed. Imagine, in the case of the function $\underline{w} = \underline{z}^{1/2}$, the complex $\underline{z}$-plane is composed of two different "sheets," as depicted in Figure 13.4. The contour that $\underline{z}$ follows in Figure 13.3 can be viewed not as the same circle twice, but one extended circle, which covers both sheets. In the $\underline{z}$-plane, as θ is increased from zero to 2π, a circle is traced out on the upper sheet in the $\underline{z}$-plane and maps to the upper half circle in the $\underline{w}$-plane,
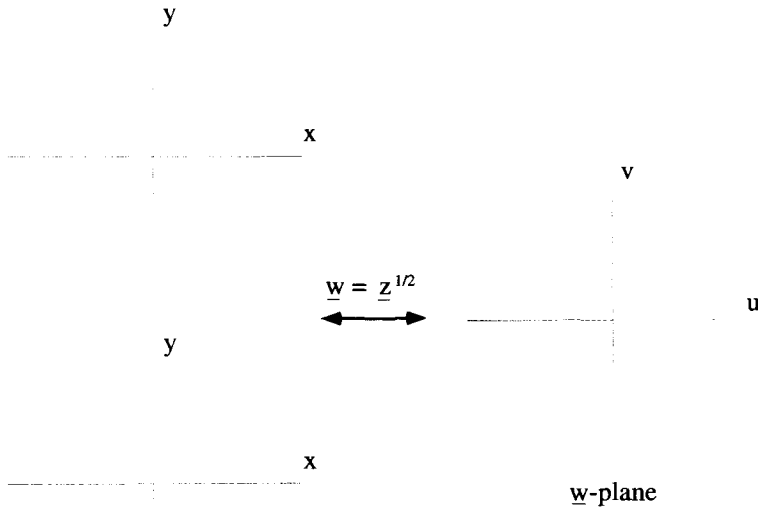
Multisheeted $\underline{z}$-plane

Figure 13.4 Multisheeted $\underline{z}$-Plane
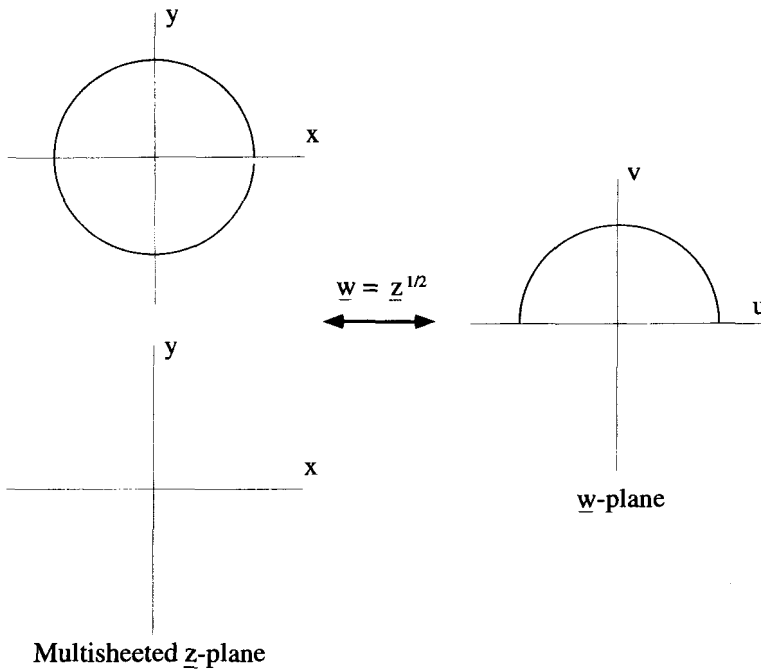
Figure 13.5   Mapping of the Unit Circle with a Multisheeted $\underline{z}$-Plane

as shown in Figure 13.5. Once θ is increased beyond 2π, suddenly $\underline{z}$ "jumps" to the second sheet in the $\underline{z}$-plane. Consequently, the $\underline{z}$-plane contour has not closed on itself at this point. As θ is increased from 2π to 4π, now tracing out a circle on the lower $\underline{z}$-plane sheet, $\underline{w}$ traces out the lower half circle in its plane, as shown in Figure 13.6. Finally, once θ reaches 4π, $\underline{z}$ jumps back to the upper sheet and the circular paths in both planes are complete.

The picture, in Figure 13.2, of how a single point gets mapped from the $\underline{z}$-plane to the $\underline{w}$-plane is also modified when we are using Riemann sheets. The point $\underline{z} = e^{i\pi/4}$ is located on the upper Riemann sheet of the $\underline{z}$-plane and maps into the point $\underline{w} = e^{i\pi/8}$ in the $\underline{w}$-plane. The point $\underline{z} = e^{i9\pi/4}$, which is located on the lower Riemann sheet, maps onto the other point $\underline{w} = e^{i9\pi/8}$ of the $\underline{w}$-plane. In this way the mapping is one-to-one, as shown in Figure 13.7.

Notice that a one-to-one mapping has been accomplished with only one sheet for the $\underline{w}$-plane. This works because the inverted function $\underline{z} = \underline{w}^2$ is not a multivalued function of $\underline{w}$. In more complicated functions, multiple Riemann sheets may be needed in both the $\underline{z}$- and $\underline{w}$-planes.

Branch Points   In the analysis of the previous section, we demonstrated the necessity of having two Riemann sheets by determining how the points along a closed contour in the $\underline{z}$-plane were mapped on the $\underline{w}$-plane. We will use this method so often

that connects the two branch points. It is for this reason that branch points must occur in pairs. They provide the end points for the branch cuts. If there is only a single branch point in the finite complex plane, it must be coupled with one at infinity. More will be said about relocating branch cuts in the examples that follow.

13.1.3   Applications of Riemann Sheet Geometry

Now that the fundamentals of branch points and branch cuts have been introduced, the details of Riemann sheet geometry are probably best understood by looking at examples of specific functions. The examples that follow discuss increasingly more complicated multivalued functions. In each case, we determine the location of the branch points and the number of Riemann sheets needed for a one-to-one mapping, locate the branch cuts, and make the Riemann sheet connections.

Example 13.3   Consider the Riemann sheets associated with the function $\underline{w} = z^{1/3}$. For this function, each value of $\underline{z}$ generates three different values of $\underline{w}$, so it's likely that there must be three Riemann sheets to describe the $\underline{z}$-plane. A tour in the $\underline{z}$ plane shows that $\underline{z} = 0$ is a branch point, and it must be circled three times before $\underline{w} = z^{1/3}$ returns to its starting value. This confirms that there must be three $\underline{z}$-plane sheets.

Figure 13.14   Riemann Sheet Connections for $\underline{w} = \underline{z}^{1/3}$

that we give it a special name. The paths used to determine the multivalued nature of a function will be referred to as *tours*. With the function $w = z^{1/2}$, we discovered a single tour around the origin returned z to its initial value, but did not return w to its starting value. Two successive tours, however, brought both z and w back to their initial values. This is why we needed two Riemann sheets in order to make our mapping one-to-one.

Still using the function $w = z^{1/2}$, imagine another tour, this time not around the origin, but around another point in the first quadrant, as shown in Figure 13.8. We label the starting point for the tour as P, with coordinates r_o and θ_o. As we move around this circle once, $z = r\,e^{i\theta}$ clearly returns to its initial value. But notice, w also returns to its starting value as well, because both r and θ have retu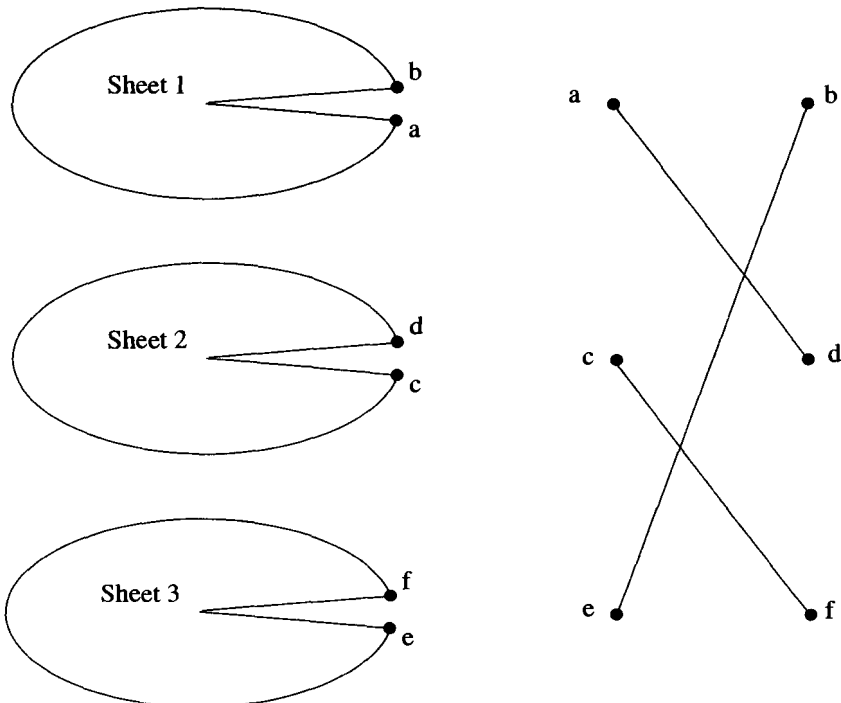rned back to their initial values. This will be the case for any tour of this function that does not surround the origin. For this particular function, the origin appears to be a very special point.

Points of this type, those around which a single tour does not return the function to its initial value, are called *branch points*. Branch points are isolated by reducing the radius of the touring circle to an arbitrary, small value. Usually the first task in making sense out of any multivalued complex integrand is to determine the locations of all its branch points.

As you will see in the next section, branch points always occur in pairs. We have seen that the function $w = z^{1/2}$ has a branch point located at the origin in the complex z-plane. Tours with this function, like the one in Figure 13.8, show that there are no other branch points of $w = z^{1/2}$ in the finite z-plane. The second branch point for this function must be located at infinity. It may seem strange to call infinity a single point, but we can do so with the aid of the interesting mapping shown in Figure 13.9. This figure shows how points in the complex plane can be mapped onto points on the surface of a sphere of unit radius. The complex plane passes through the center of the sphere. To determine the point on the sphere that corresponds to a given complex number, draw a line from N, the north pole of the sphere, to the point in the complex plane. Where this line intersects the sphere is the mapped point. Points in the complex plane that lie outside the intersection of the sphere and the plane, such as the point P_1 shown in Figure 13.9, map onto the northern hemisphere.
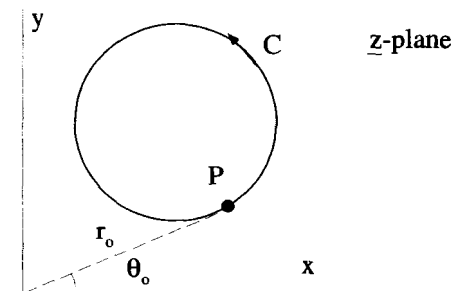
Figure 13.8   Displaced Unit Circle Contour
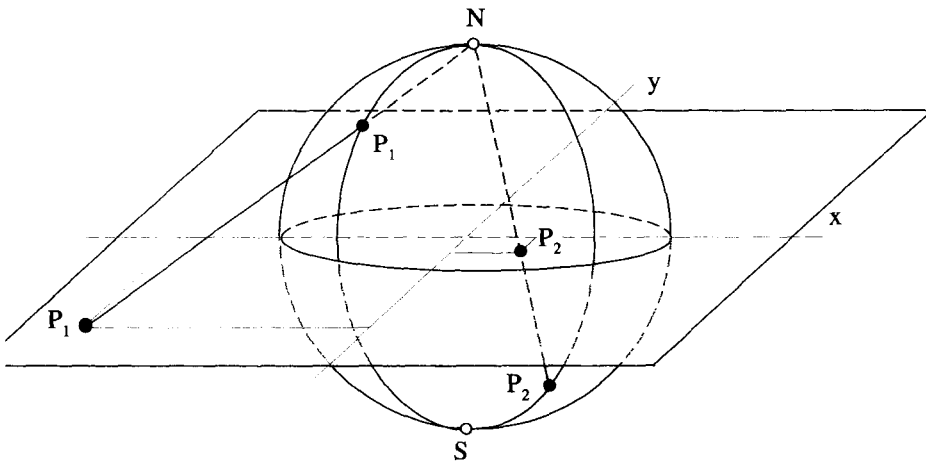
Figure 13.9   Mapping of the Complex Plane onto the Surface of a Unit Sphere

Points inside this intersection, such as P_2, map onto the southern hemisphere. Notice how the origin of the complex plane maps to the south pole of the sphere, while *any* infinity in the complex plane, that is $+\infty$, $-\infty$, $+i\infty$, $-i\infty$, etc. maps to the north pole. This single point at infinity is the second branch point of $\underline{w} = \underline{z}^{1/2}$. With this picture, a small circular tour around the origin of the $\underline{z}$-plane is equivalent to a small circular tour around the south pole. A tour around an infinitely large circle in the $\underline{z}$-plane is equivalent to a small circular tour around the north pole.

Branch Cuts   The two Riemann sheets associated with the function $\underline{z}^{1/2}$ have been described in Figures 13.4–13.7. The picture for the Riemann sheet geometry is not complete, however, until the sheets have been connected. This connection occurs where tours jump from one sheet to the other. The tour described in Figures 13.5 and 13.6 was on a circle of constant radius. As the radius of this circle is changed, it can be seen that the jump between sheets occurs along a line that falls on the positive x-axis. This line forms what is referred to as a *branch cut* and is shown in Figure 13.10. It is called a cut, because you can imagine taking a pair of scissors and actually cutting the $\underline{z}$-plane sheets along this line to prevent a contour from crossing it. It can be seen that the cut goes between the two branch points of $\underline{z}^{1/2}$. It is made on each sheet of the $\underline{z}$-plane, along the same line, as shown in Figure 13.11. The edges of the cut on one sheet have been labeled *a-b* and the edges of the other sheet have been labeled *c-d*. These edges are then connected as shown in Figure 13.12. As tours cross the branch cut they follow the connection and switch sheets.

Figure 13.12 is quite clever because it simultaneously shows the nature of the $\underline{z}$-plane from two different points of view. If the two $\underline{z}$-plane Riemann sheets are stacked and viewed from the top, the point $\underline{z} = e^{i\pi/4}$, on the top sheet, is exactly above the point $\underline{z} = e^{i9\pi/4}$, on the bottom sheet. Consequently, from this point of view these two points appear as the single point $\underline{z} = 1/\sqrt{2} + i/\sqrt{2}$. But when the sheets
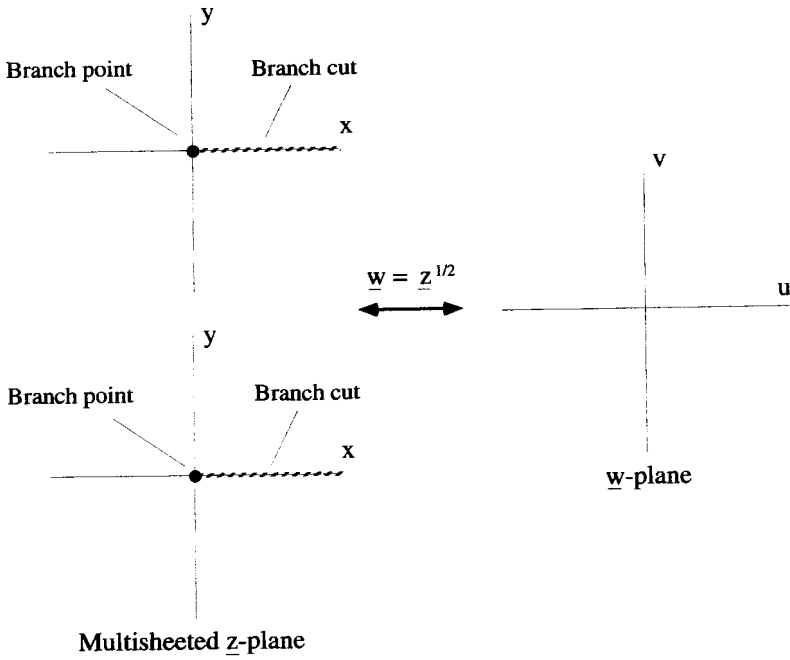
Figure 13.10   A Branch Cut in the $\underline{z}$-Plane
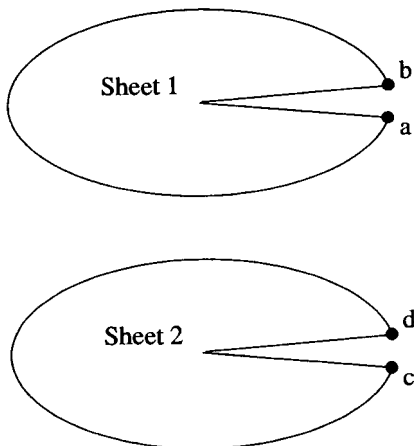
Figure 13.11   Cut Riemann Sheets
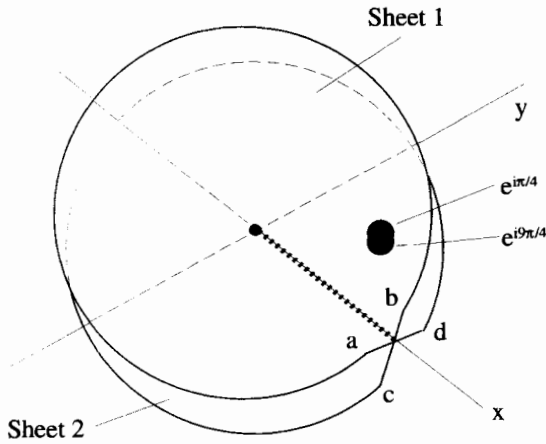
Figure 13.12   Riemann Sheet Connections

are viewed separately, $z = 1/\sqrt{2} + i/\sqrt{2}$ becomes two distinct points, one on each of the z-plane Riemann sheets. With many functions, however, making a drawing like Figure 13.12 becomes impossible. In these cases, the more abstract method of indicating the connections shown in Figure 13.13 is more convenient.

For the function $z^{1/2}$ we have placed the branch cut along the positive x-axis, between the branch points at the origin and at infinity. The arguments associated with Figures 13.5 and 13.6 led to this conclusion. With a little thought, it should be clear that the jump between sheets, and therefore the location of the branch cut for this function, does not have to be along the positive x-axis, but could be along any line
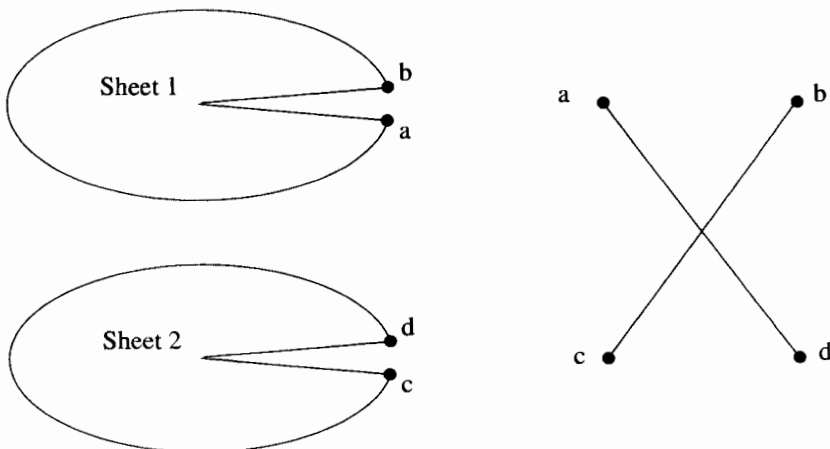
Figure 13.13   Another Way to Show Riemann Sheets Connections

that connects the two branch points. It is for this reason that branch points must occur in pairs. They provide the end points for the branch cuts. If there is only a single branch point in the finite complex plane, it must be coupled with one at infinity. More will be said about relocating branch cuts in the examples that follow.

13.1.3   Applications of Riemann Sheet Geometry

Now that the fundamentals of branch points and branch cuts have been introduced, the details of Riemann sheet geometry are probably best understood by looking at examples of specific functions. The examples that follow discuss increasingly more complicated multivalued functions. In each case, we determine the location of the branch points and the number of Riemann sheets needed for a one-to-one mapping, locate the branch cuts, and make the Riemann sheet connections.

Example 13.3   Consider the Riemann sheets associated with the function $\underline{w} = z^{1/3}$. For this function, each value of $\underline{z}$ generates three different values of $\underline{w}$, so it's likely that there must be three Riemann sheets to describe the $\underline{z}$-plane. A tour in the $\underline{z}$ plane shows that $\underline{z} = 0$ is a branch point, and it must be circled three times before $\underline{w} = z^{1/3}$ returns to its starting value. This confirms that there must be three $\underline{z}$-plane sheets.

Figure 13.14   Riemann Sheet Connections for $\underline{w} = \underline{z}^{1/3}$

This function is so similar to $\underline{w} = \underline{z}^{1/2}$, that it is fairly clear that the only other branch point is at infinity. There is a single branch cut connecting these two branch points. If, as before, the cut is made along the positive real axis, the connections of the Riemann sheets are as shown in Figure 13.14. If you start on the top sheet (sheet #1), a single loop around the origin drops you to the second sheet. A second tour drops you to the third sheet. A third loop returns you to the first sheet, and back to your starting value of $\underline{w}$. This example points out the usefulness of the simple dot connection diagram for indicating sheet connections. It would be very difficult to make a drawing like Figure 13.12 for this function.

Example 13.4 As the next example, consider the function $\underline{w} = \ln \underline{z}$. This function is more conveniently handled with polar representation: $\underline{z} = re^{i\theta}$ and $\underline{w} = \ln r + i\theta$. From this form, it is clear that all tours in the $\underline{z}$-plane around any point other than the origin will return $\underline{w}$ to its initial value. A tour around the origin, however, will not. To see this, start a tour around the origin at some initial point with $r = r_i$ and $\theta = \theta_i$. The initial value of the function is $\underline{w} = \ln r_i + i\theta_i$. After a single lap around the origin, $\underline{w} = \ln r_i + i(\theta_i + 2\pi)$. After two laps, $\underline{w} = \ln r_i + i(\theta_i + 4\pi)$. No matter how many complete tours are made, $\underline{w} = \ln \underline{z}$ will never return to its initial value! Consequently, for this function, there are an infinite number of $\underline{z}$-plane Riemann sheets. If the sheets are cut along the positive real axis, they are connected as shown in Figure 13.15.

Example 13.5 Now consider the function

$$\underline{w} = (\underline{z}^2 - a^2)^{1/2}, \tag{13.19}$$

where a is a positive real number. This example is quite a bit more complicated, and requires the use of multiple complex phasors to check for branch points. One way to approach a function of this nature would be to break it up into two or more simple intermediate functions. For instance, in this case you might let $\underline{t} \equiv \underline{z}^2 - a^2$ so $\underline{w} = \underline{t}^{1/2}$. Then you would analyze the two functions separately and combine the results. For this example, however, it is more instructive to work with the entire function directly. It does help to factor Equation 13.19 into two terms:

$$\underline{w} = (\underline{z} - a)^{1/2}(\underline{z} + a)^{1/2}. \tag{13.20}$$

Our first task is to find all the branch points in the $\underline{z}$-plane. To accomplish this, we must perform lots of tours to determine where $\underline{w}$ does not return to its initial value. It is useful to define two phasors for the complex quantities $\underline{z} - a$ and $\underline{z} + a$, as shown in Figure 13.16. These phasors are, in exponential notation,

$$\underline{z} + a = r_1 e^{i\theta_1}$$

$$\underline{z} - a = r_2 e^{i\theta_2}. \tag{13.21}$$

With these definitions, $\underline{w}$ can be written as

$$\underline{w} = \left(r_1 e^{i\theta_1}\right)^{1/2} \left(r_2 e^{i\theta_2}\right)^{1/2} \tag{13.22}$$
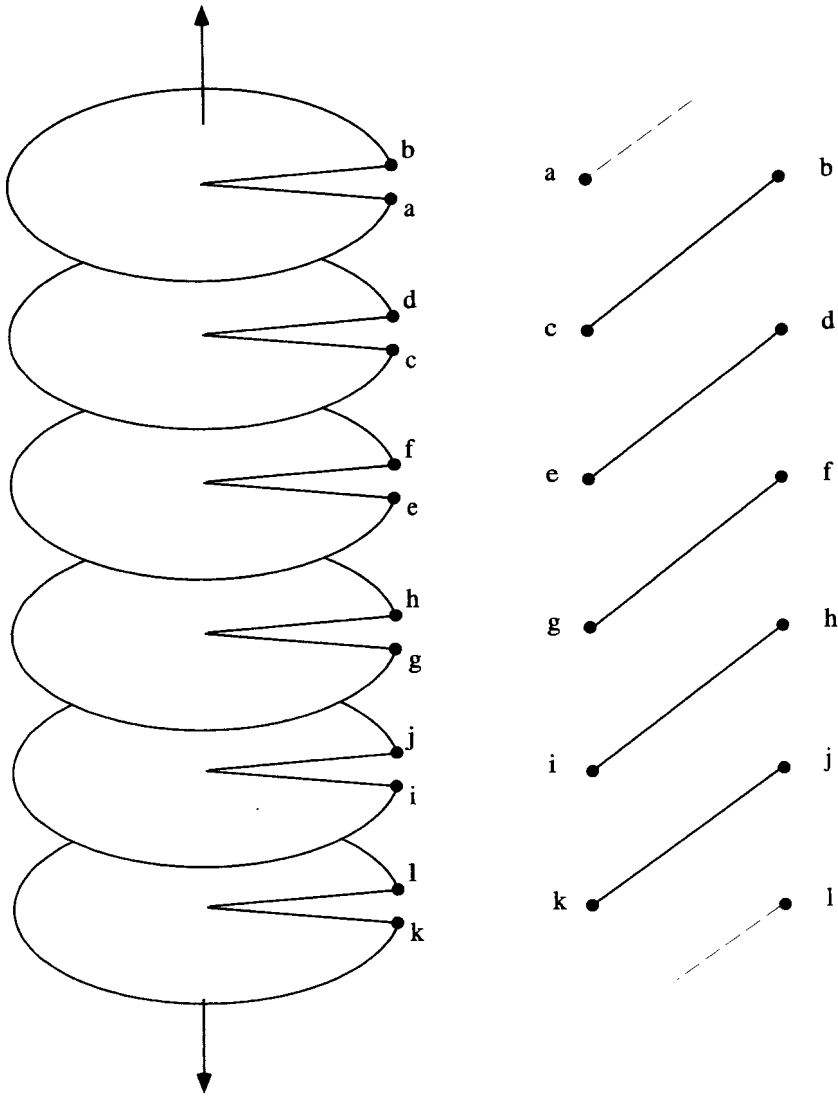
Figure 13.15 Riemann Sheets for $\underline{w} = ln\,\underline{z}$
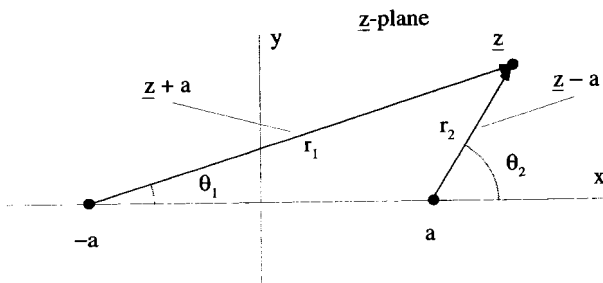
Figure 13.16 Phasors for Touring in the z-Plane
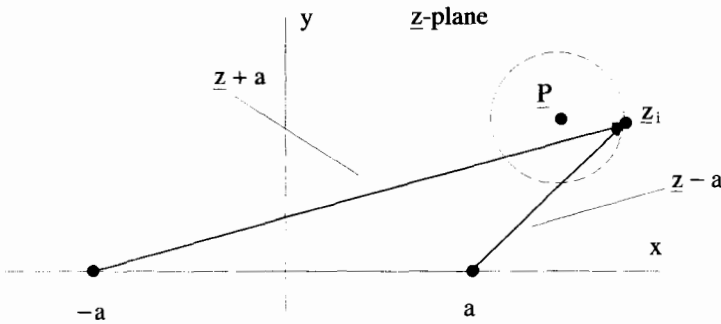
Figure 13.17 Tour Around $z \neq \pm a$

or

$$\underline{w} = \sqrt{r_1}\sqrt{r_2}\,e^{i(\theta_1 + \theta_2)/2}. \tag{13.23}$$

For the first tour, pick any point P, which is not one of the two points $\pm a$, as shown in Figure 13.17. At the start of the tour, $\underline{z} = \underline{z}_i$, with the corresponding initial values for the phasor components being r_{1i}, θ_{1i}, r_{2i}, and θ_{2i}. A single counterclockwise tour around P returns both $\underline{z}$ and all these phasor components to their initial values. Consequently, $\underline{w}$ also returns to its initial value, and P is not a branch point of $\underline{w} = (\underline{z}^2 - a^2)^{1/2}$.

Now let's tour around the point $\underline{z} = a$, again defining the starting position of the tour to be $\underline{z} = \underline{z}_i$, as shown in Figure 13.18. The initial values for the phasors are again r_{1i}, θ_{1i}, r_{2i}, and θ_{2i}. As $\underline{z}$ makes a counterclockwise loop around this point and returns to its initial value, r_1, θ_1, and r_2 all return to their initial values, but this time θ_2 has increased to $\theta_{2i} + 2\pi$. Consequently, $\underline{w}$ after one loop is

$$\begin{aligned}
\underline{w} &= \sqrt{r_{1i}}\sqrt{r_{2i}}\,e^{i(\theta_{1i} + \theta_{2i} + 2\pi)/2} \\
&= -\sqrt{r_{1i}}\sqrt{r_{2i}}\,e^{i(\theta_{1i} + \theta_{2i})/2},
\end{aligned} \tag{13.24}$$
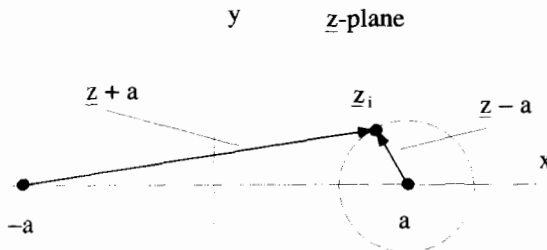
Figure 13.18 Tour Around the Point $\underline{z} = a$

which is a factor of -1 different from its initial value. This shows $\underline{z} = a$ is a branch point of $\underline{w} = (\underline{z}^2 - a^2)^{1/2}$. A tour around the point $\underline{z} = -a$ produces a similar result, so the two points $\underline{z} = \pm a$ are both branch points of $\underline{w}$.

We have already shown there are no other branch points in the finite $\underline{z}$-plane, but there is the possibility that these two branch points do not form a pair with a single cut between them. Instead, each of the branch points might have a separate cut extending out to infinity. In this case, you can visualize each of the two branch points paired with a different branch point at infinity. An easy way to determine if this is actually happening in this case is to make a large tour that encloses both branch points, as shown in Figure 13.19. If, at the end of this tour, we are on a different sheet from the start, we must have crossed some cut(s) extending to infinity. After performing this big tour, notice both r_1 and r_2 have returned to their initial values, but *both* θ_1 and θ_2 have increased by 2π, so $\underline{w}$ becomes:

$$
\begin{aligned}
\underline{w} &= \sqrt{r_{1i}} \sqrt{r_{2i}}\, e^{i(\theta_{1i} + 2\pi + \theta_{2i} + 2\pi)/2} \\
&= \sqrt{r_{1i}} \sqrt{r_{2i}}\, e^{i(\theta_{1i} + \theta_{2i})/2},
\end{aligned}
\tag{13.25}
$$

which is exactly the initial value of $\underline{w}$. We did not have to cross a cut, so the two branch points at $\underline{z} = \pm a$ must form a pair with a single cut between them.

To determine the complete nature of the sheets we need to know the results of a few more tours. It is easy to show that touring around either of the branch points at $\underline{z} = \pm a$ twice returns $\underline{w}$ to its initial value. Also, going around one of the branch
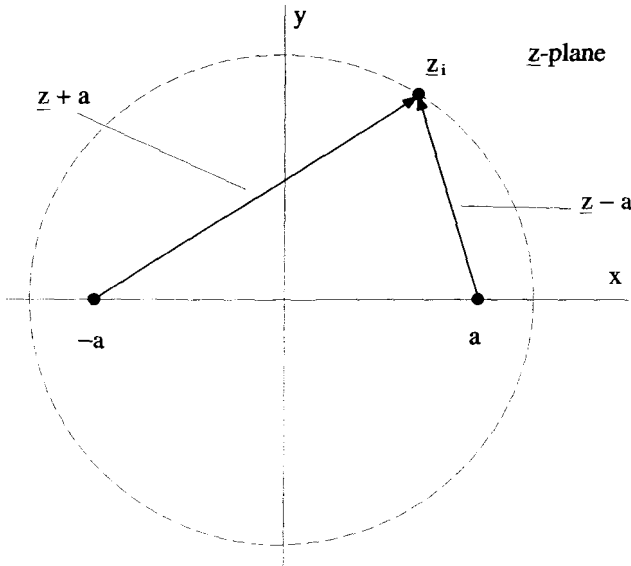
Figure 13.19   Tour Around Both Branch Points
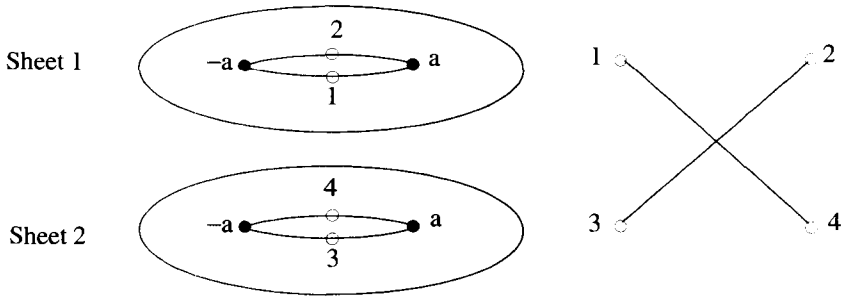
Figure 13.20 Riemann Sheet Connections

points once and then back around the other, in a figure-eight pattern, also returns $\underline{w}$ to its initial value. These facts lead us to conclude there are two Riemann sheets in the $\underline{z}$-plane, and they are connected with a cut between them, as shown in Figure 13.20.

This figure shows the cut placed along a straight line between $\underline{z} = \pm a$. As we mentioned earlier, the cut can be placed anywhere as long as it terminates at the appropriate branch points. We could legitimately deform the cut, in successive steps, as shown in Figure 13.21. The configuration of the cut at the bottom of this figure makes it look like there might now be two branch cuts. This is not the case. It is actually the same cut making a very large loop around the complex plane. With the cut positioned as shown at the bottom of Figure 13.21, a tour enclosing both branch points crosses the branch cut twice. Because this is the same cut, however, the second crossing of the cut returns us to the original sheet.

Example 13.6 As a last example, consider the function

$$\underline{w} = (\underline{z} - a)^{1/2} + (\underline{z} + a)^{1/2}. \tag{13.26}$$

Although this looks very similar to the function in the previous example, it behaves quite differently. To find the branch points, tours must be made in the $\underline{z}$-plane. The same phasors that were used in the previous problem, and shown in Figure 13.16, can be used again here:

$$\underline{z} + a = r_1 e^{i\theta_1}$$
$$\underline{z} - a = r_2 e^{i\theta_2}. \tag{13.27}$$

With these phasors, the function can be rewritten

$$\underline{w} = \sqrt{r_1}\, e^{i\theta_1/2} + \sqrt{r_2}\, e^{i\theta_2/2}. \tag{13.28}$$

In a single, counterclockwise lap around $\underline{z} = a$, r_1, θ_1 and r_2 return to their initial values. The angle θ_2, however, increases by 2π. For simplicity, take the initial values of θ_1 and θ_2 to be zero, and the initial radial values to be $r_1 = r_{1i}$ and $r_2 = r_{2i}$. Then
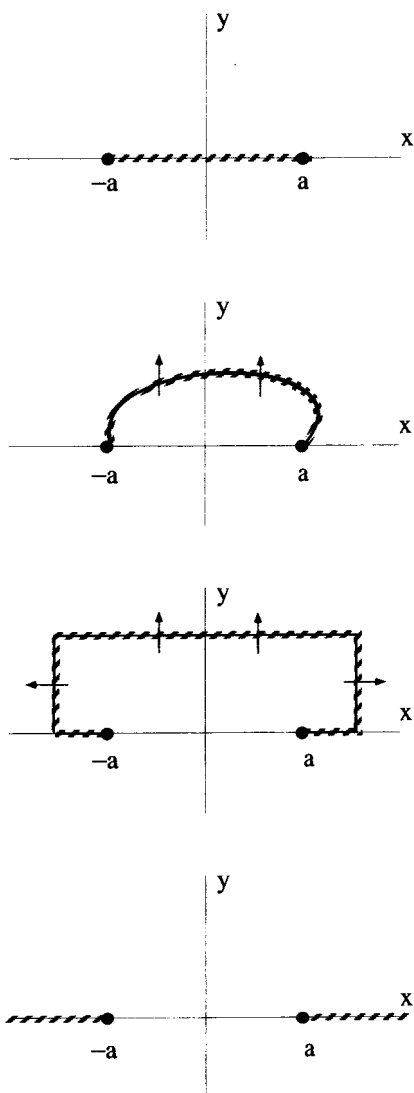
Figure 13.21   Deformation of the Branch Cut

at the start of this tour,

$$\underline{w} = \sqrt{r_{1i}} + \sqrt{r_{2i}}, \tag{13.29}$$

but at the end

$$\underline{w} = \sqrt{r_{1i}} - \sqrt{r_{2i}}. \tag{13.30}$$

A tour around $z = -a$ produces a similar result, so both points at $z = \pm a$ are branch points. Single loop tours around any other point in the finite z-plane will return $\underline{w}$ to its starting value, so there are no other branch points in the finite z-plane.

As was the case in the previous example, we must check to see if there are any branch points at infinity by making a tour around both $z = a$ and $z = -a$. Start the tour with both θ_1 and θ_2 equal to zero. At the end of this tour, r_1 and r_2 will return to their initial values, but both θ_1 and θ_2 will increase by 2π. Thus, at the start of this tour

$$\underline{w} = \sqrt{r_{1i}} + \sqrt{r_{2i}}, \tag{13.31}$$

while at the end

$$\underline{w} = -\sqrt{r_{1i}} - \sqrt{r_{2i}}. \tag{13.32}$$

Consequently, there must be two branch points at infinity, and the branch cuts that start at $z = a$ and $z = -a$ extend to infinity, as shown in Figure 13.22. Unlike the single cut, with the misleading appearance of two cuts, shown at the bottom of Figure 13.21, these really are two distinct cuts extending out to infinity.

We can determine the total number of Riemann sheets and how they are connected at the cuts by performing a few more tours, and watching how many different values of $\underline{w}$ are generated at the same point of z. Consider the three different tours shown in Figure 13.23. These tours all start with the same value of z. This value, which we will call $\underline{z_i} = r_i\, e^{i\theta_i}$, is a pure real quantity, lying just above the cut with $\theta_i = 0$ and to the right of $z = a$. In terms of the phasor parameters, at the start of each of these tours

$$\underline{w} = \sqrt{r_{1i}} + \sqrt{r_{2i}}. \tag{13.33}$$

We already determined that at the end of the topmost tour

$$\underline{w} = \sqrt{r_{1i}} - \sqrt{r_{2i}}, \tag{13.34}$$
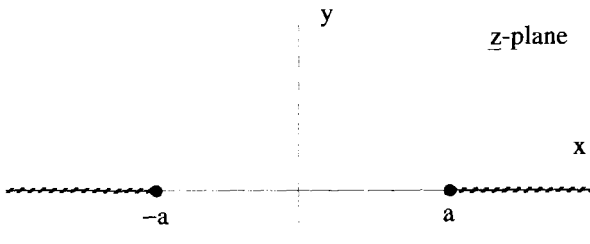
Figure 13.22 Branch Points and Branch Cuts for $\underline{w} = (z - a)^{1/2} + (z + a)^{1/2}$
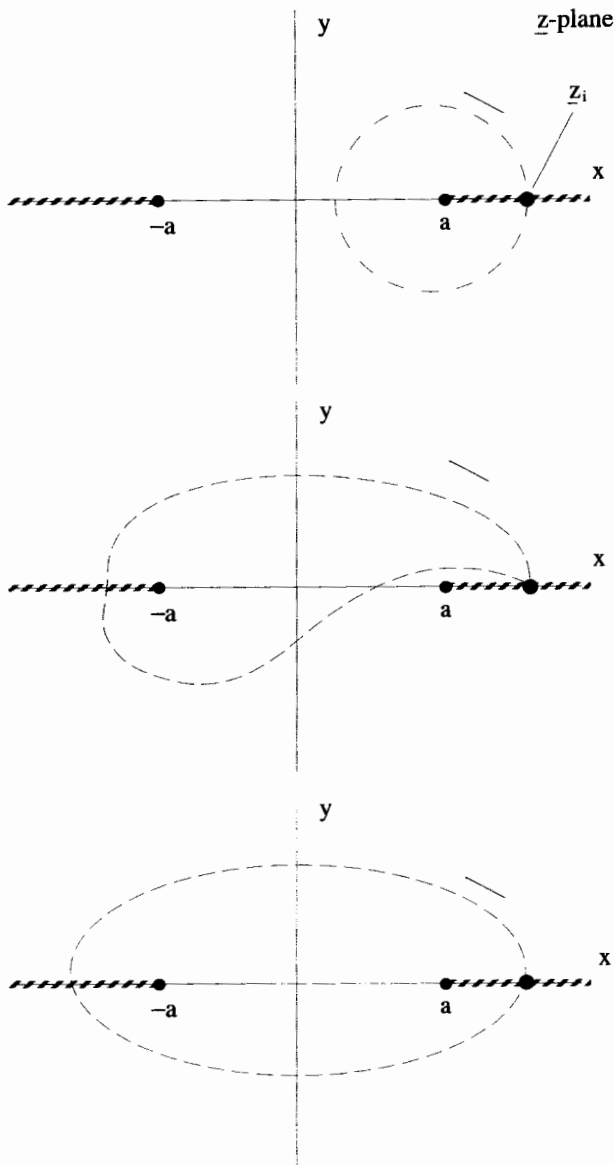
Figure 13.23   Tours to Determine the Number of Riemann Sheets

which is different from the initial value, so at the end of this tour we must be on a
second Riemann sheet. Now consider the tour shown in the middle of Figure 13.23.
After one lap around this path, only θ_1 has changed by 2π, and so

$$\underline{w} = -\sqrt{r_{1i}} + \sqrt{r_{2i}}. \qquad (13.35)$$

This is a third value for $\underline{w}$, different from both Equation 13.34 and 13.35, so at the end of this tour, we must be on a third Riemann sheet. A single lap around the tour shown on the bottom of Figure 13.23 changes both θ_1 and θ_2 by 2π. At the end of this tour

$$\underline{w} = -\sqrt{r_{1i}} - \sqrt{r_{2i}}. \tag{13.36}$$

This is a fourth value for $\underline{w}$ and consequently requires a fourth Riemann sheet.

Any combination of these tours or multiple laps of a single tour will not generate any new values for $\underline{w}$, and so we conclude that there are four Riemann sheets. Because a two-lap tour around either of the branch points at $\underline{z} = \pm a$ would return us to the original sheet, but a figure-eight pattern would not, the connection diagram must be as shown in Figure 13.24.
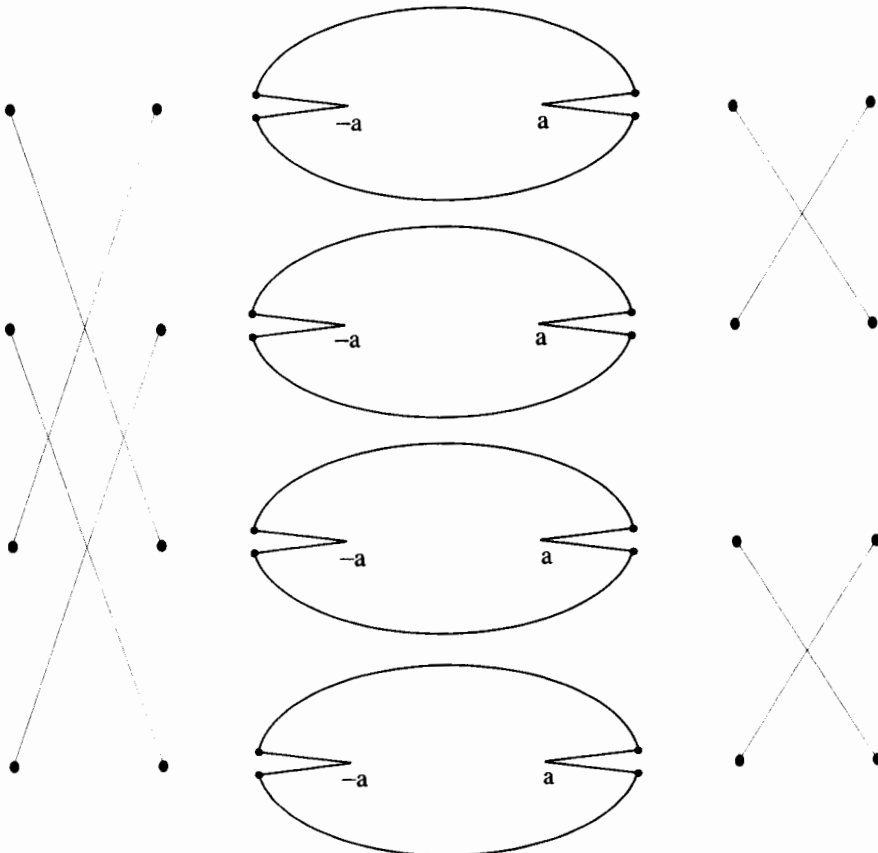
Figure 13.24 Riemann Sheet Connections

A quicker way to see that there should be four Riemann sheets for this function is to realize that the $\pm$ property of the square root operation, i.e.,

$$\underline{w} = \pm\sqrt{r_1} \pm \sqrt{r_2}, \tag{13.37}$$

allows four different possible combinations for forming $\underline{w}$. This method is useful for determining the number of sheets, but does not guide you in determining how they are connected. In our previous example, with $\underline{w} = (\underline{z} - a)^{1/2}(\underline{z} + 2)^{1/2}$, the square roots have only two combinations for $\underline{w}$,

$$\underline{w} = \pm\sqrt{r_1}\sqrt{r_2}, \tag{13.38}$$

and so there were only two Riemann sheets associated with this function.

13.1.4   Riemann Sheet Mapping

Quite often, it is useful to see how an entire Riemann sheet gets mapped from one plane onto another. This is quite easy for the function $\underline{w} = \underline{z}^{1/2}$. If we define $\underline{z} = r_z e^{i\theta}$ and $\underline{w} = r_w e^{i\phi}$, we can write

$$r_w = \sqrt{r_z} \tag{13.39}$$

$$\phi = \frac{\theta}{2}. \tag{13.40}$$

According to Figure 13.10, we are on the first Riemann sheet if $0 < \theta < 2\pi$. Therefore, the values of $\underline{w}$ we can generate from the first sheet will be limited to the angular range $0 < \phi < \pi$, or equivalently the upper half plane. Likewise, the second Riemann sheet will map onto the lower half plane of the $\underline{w}$-plane. This situation is shown in Figure 13.25.

Now remember, we said the actual placement of the branch cut is arbitrary, as long as it connects the branch points. For this function, the branch cut could equally well have been placed along the positive imaginary axis of the $\underline{z}$-plane. In that case, the $\underline{z}$-plane sheets would map onto the $\underline{w}$-plane as shown in Figure 13.26. Now the first Riemann sheet is associated with the range $\pi/2 < \theta < 5\pi/2$, and the second sheet with $5\pi/2 < \theta < 9\pi/2$. In the $\underline{w}$ plane, the dividing line has been rotated by $\pi/4$. Keep in mind, there is nothing that says the branch cut has to be a straight line. It could be placed as shown in Figure 13.27. In this case, it is impossible to define the different Riemann sheets of the $\underline{z}$-plane using a simple range of θ. Don't make the common mistake of assuming a Riemann sheet is always defined by an angular range.

It is possible that, in order to get a one-to-one mapping, there must be multiple sheets in both the $\underline{z}$- and $\underline{w}$- planes. Consider the function $\underline{w} = \underline{z}^{3/2}$. We investigated this function briefly at the beginning of this chapter. We showed there were two values of $\underline{w}$ for a single value of $\underline{z}$, so there must be two $\underline{z}$-plane Riemann sheets.
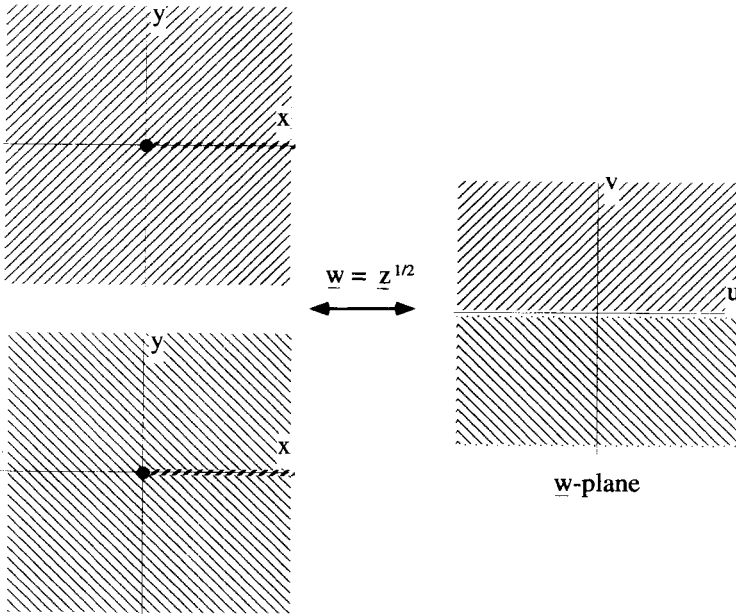
Multisheeted $\underline{z}$-plane

Figure 13.25   Riemann Sheet Mapping for $\underline{w} = \underline{z}^{1/2}$
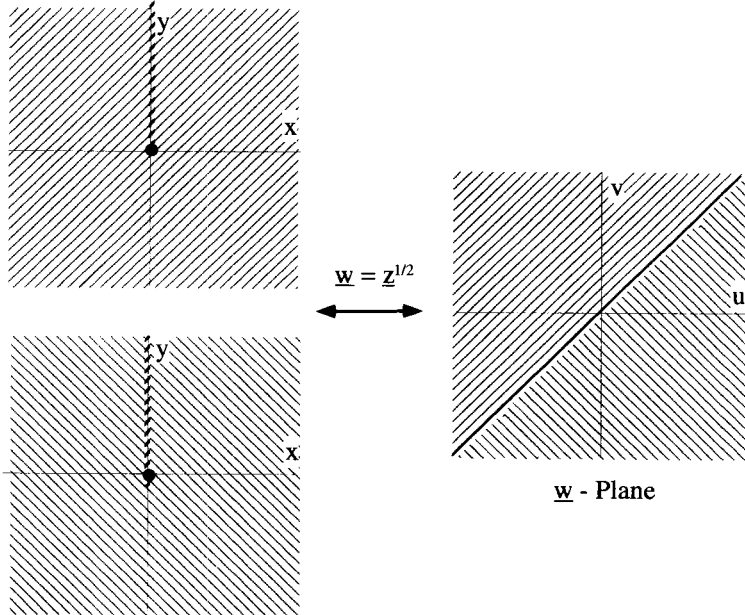
Multisheeted $\underline{z}$ - Plane

Figure 13.26   Another Riemann Sheet Mapping for $\underline{w} = \underline{z}^{1/2}$
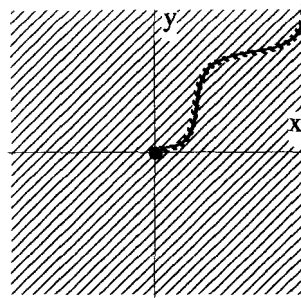
Top sheet of the $\underline{z}$-plane

Figure 13.27   Arbitrary-Shaped Branch Cut

Quite similar to the square root function, tours in the $\underline{z}$-plane reveal two branch points, one at $\underline{z} = 0$ and one at $\underline{z} = \infty$. In this case, however, the inversion of this function $\underline{z} = \underline{w}^{2/3}$ is also multivalued. Each value of $\underline{w}$ produces three different values of $\underline{z}$, so there are three Riemann sheets in the $\underline{w}$-plane. Tours in the $\underline{w}$-plane show that there are two branch points, one at $\underline{w} = 0$ and one at $\underline{w} = \infty$. If the cuts in both planes are taken along the positive, real axes, the $\underline{z}$-plane sheets map onto the $\underline{w}$-plane sheets, as shown in Figure 13.28. The branch cut in the $\underline{z}$-plane maps onto the real axis of the second Riemann sheet of the $\underline{w}$-plane.

13.1.5   Integrals of Multivalued Functions

Remember, the primary reason we have spent all this effort defining branch points, branch cuts, and Riemann sheets is so we can use the Residue theorem and other Cauchy techniques to perform integrals on multivalued complex functions.

Consider a contour integral in the complex $\underline{z}$-plane of our favorite multivalued function:

$$\int_C d\underline{z}\,\underline{z}^{1/2}. \tag{13.41}$$

From our previous experience with this function, we know there are two Riemann sheets, and two branch points in the $\underline{z}$-plane, one at $\underline{z} = 0$ and the other at $\underline{z} = \infty$. For purposes of this discussion, let C extend across the lower half of the first sheet of the $\underline{z}$-plane, so that the contour looks as shown in Figure 13.29. The integrand in Equation 13.41 is analytic everywhere except at $\underline{z} = 0$, and so the contour C can be deformed without changing the value of the integral, as long as it is not moved across $\underline{z} = 0$. Let's attempt to move it up, into the upper half plane. As the contour is moved up it is snagged by the nonanalytic point at $\underline{z} = 0$. But that is not all that happens. The left half of the contour, along which the real part of $\underline{z}$ is less than zero, moves up with no problem. But as the right half of this contour is moved across the positive real
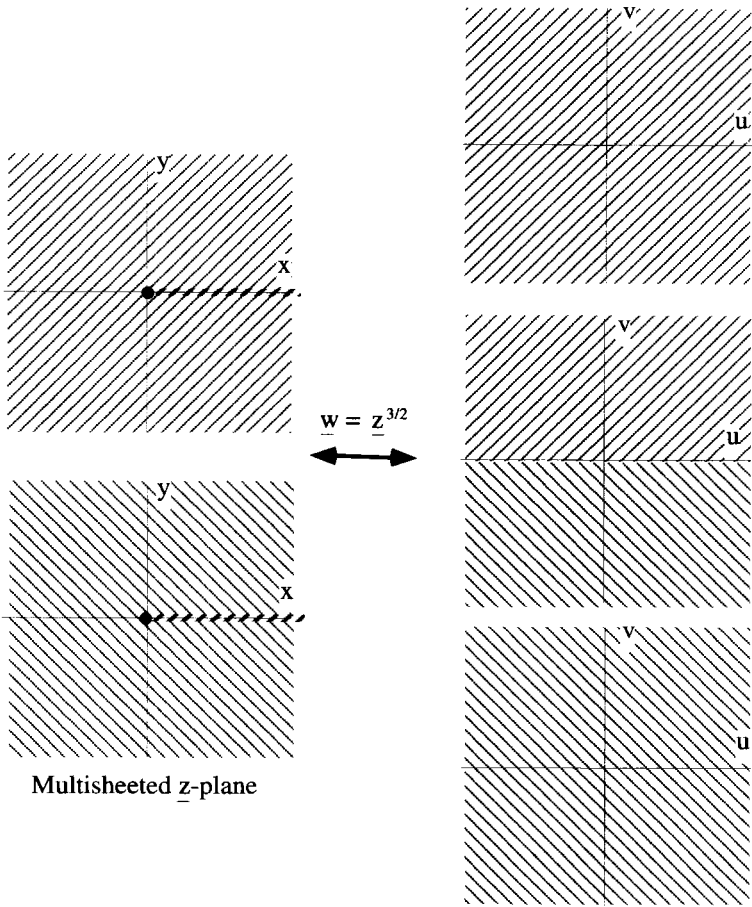
Figure 13.28 Mapping of Riemann Sheets for $\underline{w} = \underline{z}^{3/2}$
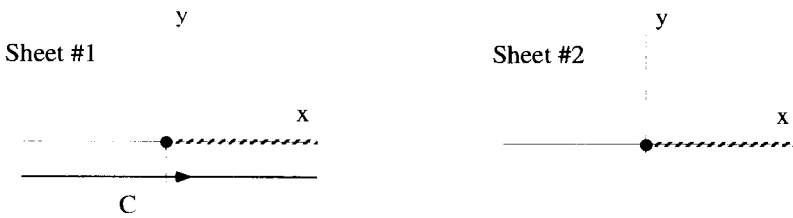
Figure 13.29 An Integration in a Multisheeted z-Plane
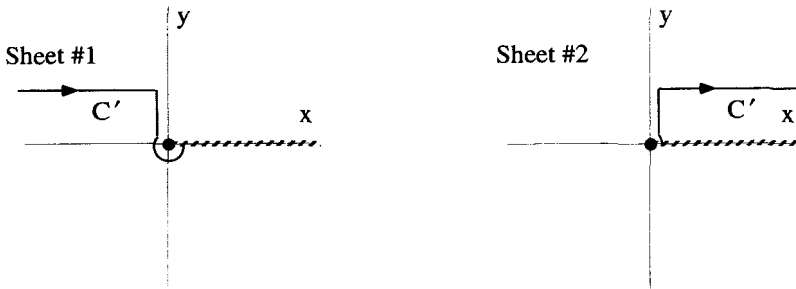
Figure 13.30   Contour Deformation with Two Riemann Sheets

axis, the deformed contour runs into the cut and must switch to the second Riemann sheet, as shown in Figure 13.30.

Actually, it is possible to deform C and stay on a single Riemann sheet. Because the location of the branch cut is arbitrary, as long as it still connects the branch points we can modify it to extend along the positive imaginary axis, as shown in Figure 13.31. Now as the contour C is raised into the upper half z-plane, it is deformed as shown in Figure 13.32.
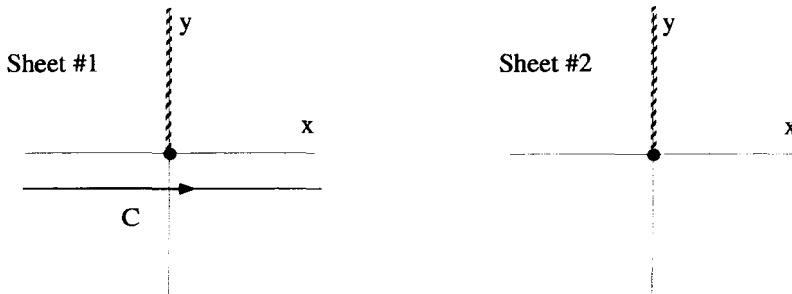
Figure 13.31   Integration with Repositioned Branch Cut
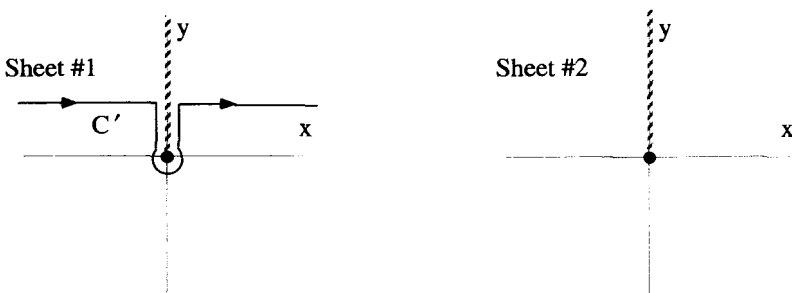
Figure 13.32   Contour Deformation Remaining on a Single Riemann Sheet

Sheet #1

$\theta = -3\pi/2$

$\theta = \pi/2$

y

x
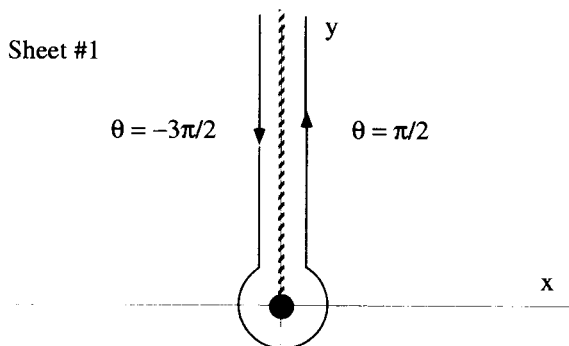
Figure 13.33 Integration Along the Branch Cut and Around the Branch Point

Continuing with this second approach to the deformation, consider the portion of this integration shown in detail in Figure 13.33. It is tempting to claim that the two vertical portions of this contour can be made to cancel each other, because they follow the same line, but in opposite directions. This would be the case if it weren't for the branch cut, which causes the function $\underline{z}^{1/2}$ to have different values on either side. To see this, let

$$\underline{z} = r\, e^{i\theta}, \tag{13.42}$$

with $\theta = 0$ being on sheet #1. Along the vertical segment just to the right of the branch cut, $\theta = \pi/2$ and

$$\underline{z}^{1/2} = \sqrt{r}\, e^{i\pi/4} = \sqrt{r/2}\,(1 + i) \qquad 0 < r < \infty. \tag{13.43}$$

Along the vertical segment just to the left of the branch cut, $\theta = -3\pi/2$ and we have

$$\underline{z}^{1/2} = \sqrt{r}\, e^{-i3\pi/4} = -\sqrt{r/2}\,(1 + i) \qquad \infty > r > 0. \tag{13.44}$$

The value of $\underline{z}^{1/2}$ on the left-hand side of the cut is not equal to its value on the right-hand side, and so the integrals along the vertical sections do not cancel.

We should also look carefully at little circular part of the integral around the branch point in Figure 13.33. To perform this integration define $\underline{z} = \epsilon e^{i\theta}$. Then as the circular contour is shrunk around the branch point,

$$\underline{z}^{1/2} = \lim_{\epsilon \to 0} \sqrt{\epsilon}\, e^{i\theta/2} \qquad -3\pi/2 < \theta < \pi/2. \tag{13.45}$$

Consequently, as $\epsilon \to 0$, the integrand also goes to zero and the contribution from this circular segment is zero. In general, however, the integral around a branch point will not be zero and its contribution must be carefully evaluated for each particular situation.

These ideas are developed in the two examples that follow.
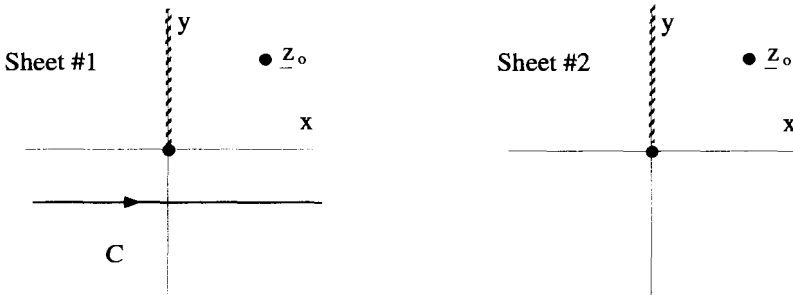
Figure 13.34   Complex Plane Picture for Integration of Example

---

Example 13.7   As an example of how contours involving multivalued functions can be deformed in the complex plane, consider the integration

$$\int_C d\underline{z}\, \frac{\underline{z}^{1/2}}{(\underline{z} - \underline{z}_o)^2}, \tag{13.46}$$

where $\underline{z}_o$ is some point in the complex plane, and C is the straight line contour on the first sheet, shown in Figure 13.34. The multivaluedness of the integrand is controlled by $\underline{z}^{1/2}$ in the numerator. The integrand can, as before, be represented on two Riemann sheets with branch points at $\underline{z} = 0$ and $\underline{z} = \infty$, and a single branch cut between them. The addition of the term in the denominator does not change the arrangement of branch points or the nature of the Riemann sheets, but it does add a nonanalytic, second-order pole on both sheets, at $\underline{z} = \underline{z}_o$. The branch cut is taken along the positive imaginary axis in anticipation of the contour deformation we are about to perform.

A clever transformation of this integral can be arrived at by deforming C into the upper half plane as far as possible, while trying to stay entirely on the first Riemann sheet. As we do this, the contour "snags" on both the branch point and the pole at $\underline{z}_o$. This situation is shown in Figure 13.35. As the imaginary part of the horizontal sections of C' approach $+i\infty$, their contribution to the integral vanishes, because the
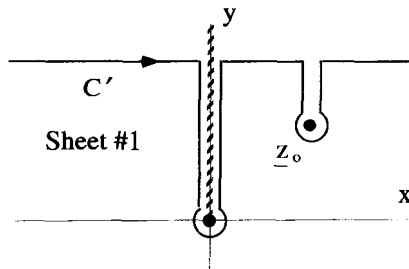
Figure 13.35   Deformed Integration Contour
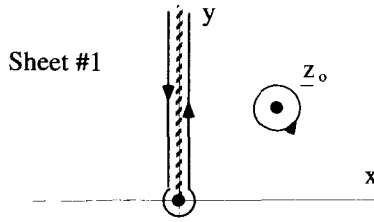
Figure 13.36   Final Contours for Evaluating Integration

integrand is approaching zero at a rate proportional to $r^{-3/2}$. The two vertical sections extending down to z_o also cancel, because the integrand has the same value on these segments and they are in opposite directions. From our previous discussion, however, we know the integration along the vertical sections on either side of the branch cut do not cancel, and the contribution around the branch point must be evaluated carefully. The contour shown in Figure 13.34 is thus equivalent to the sum of the two contours shown in Figure 13.36.

Example 13.8   Now let's work a specific, practical problem. Consider the evaluation of the real integral

$$I = \int_0^\infty dx \, \frac{1}{\sqrt{x}\,(x+1)}. \tag{13.47}$$

Remember our convention that the $\sqrt{}$ symbol always gives us the positive root, so we should obtain a positive result for this integral.

Our first step is to convert this integral into one in the complex plane by substituting $\underline{z}$ for x, and then determining an appropriate contour. The integral can be written as

$$I = \int_C d\underline{z} \, \frac{1}{\underline{z}^{1/2}(\underline{z}+1)}. \tag{13.48}$$

The integrand is multivalued because of the $\underline{z}^{1/2}$ term, and so there are two Riemann sheets, the branch points at zero and infinity, and the single branch cut between them. We place the branch cut on the positive real axis. To properly correspond to the real integral, the complex contour C must also run along the real axis, directly parallel to the cut, on one of the two sheets. To determine which one, let

$$x \to \underline{z} = r\,e^{i\theta}. \tag{13.49}$$

Along C, we must have either $\theta = 0$ (on the first sheet) or $\theta = 2\pi$ (the second sheet). But notice if $\theta = 2\pi$, we get the negative root in the denominator, which we explicitly said we did not want. Thus we take the contour to be on the first sheet, as shown in Figure 13.37.
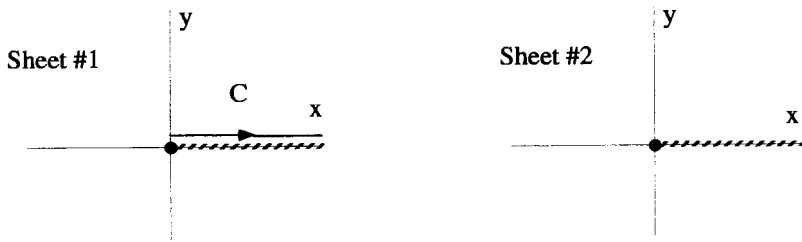
Figure 13.37 Starting Contour for Integral Evaluation

We are left with evaluating the integral in Equation 13.48 along the path shown in Figure 13.37. This integration can be evaluated by considering integration around the cleverly closed contour C_o on sheet #1, shown in Figure 13.38. This contour is indeed closed, because we have made a special effort to stay on the first sheet by avoiding the branch cut entirely. The integration around C_o is easy to evaluate, because the integrand is analytic everywhere inside C_o except at $\underline{z} = -1$. Therefore,

$$\oint_{C_o} d\underline{z} \, \frac{1}{\underline{z}^{1/2}(\underline{z}+1)} = 2\pi i \, \frac{1}{\underline{z}^{1/2}}\bigg|_{\underline{z}=e^{i\pi}} = 2\pi. \qquad (13.50)$$
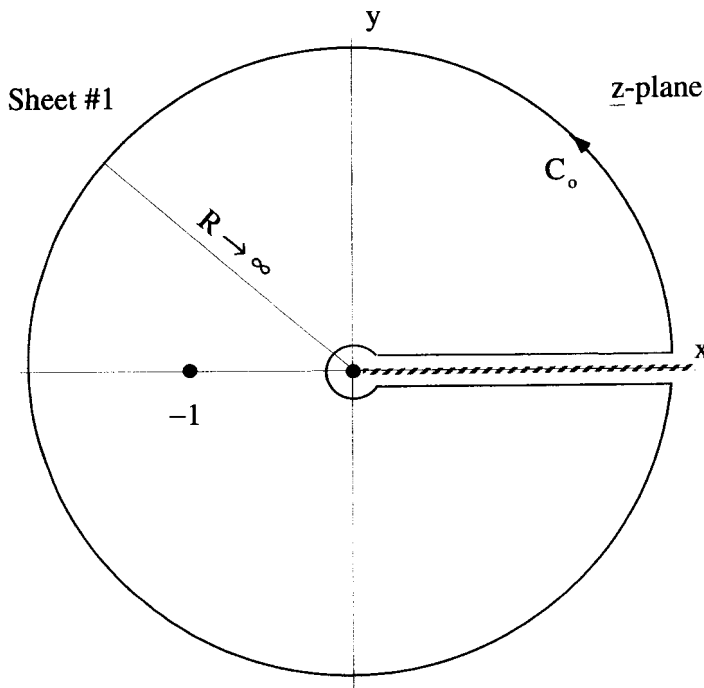
Figure 13.38 Equivalent Closed Contour Integration
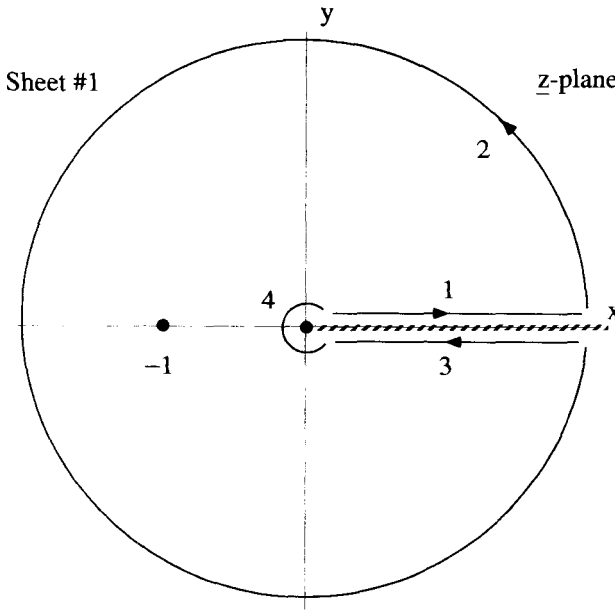
Figure 13.39   The Four Parts of the C_o Contour

It is important to notice that the residue was evaluated at $\underline{z} = e^{i\pi}$ and not at $\underline{z} = e^{i3\pi}$, because the integration is being performed on sheet #1.

The problem now becomes one of relating the outcome of this integration back to the original integral of Equation 13.48. The integration around C_o can be broken up into four parts:

$$\oint_{C_o} = \int_1 + \int_2 + \int_3 + \int_4, \tag{13.51}$$

where the paths 1-4 are shown in Figure 13.39. Along path 1, $\theta = 0$, $\underline{z} = re^{i0}$, and $d\underline{z} = dr$ so

$$\int_1 = \int_0^\infty dr \, \frac{1}{\sqrt{r} \, (r + 1)} = I. \tag{13.52}$$

The integral along path 1 is the same integral as in Equation 13.47, the very thing we hope to determine. Along path 2, we are moving around a circle of constant radius R, so $\underline{z} = Re^{i\theta}$ and $d\underline{z} = iRe^{i\theta}d\theta$, with θ ranging from 0 to 2π. As we take the limit as R goes to infinity, path 2 contributes

$$\int_2 = \lim_{R \to \infty} \int_0^{2\pi} d\theta \frac{iRe^{i\theta}}{\sqrt{Re^{i\theta/2}}(Re^{i\theta} + 1)}$$
$$= 0. \tag{13.53}$$

Along path 3, $\theta = 2\pi$, while r ranges from infinity to zero, so $\underline{z} = re^{i2\pi}$ and $d\underline{z} = re^{i2\pi}dr$. The integral on path 3 becomes

$$\int_3 = \int_\infty^0 dr\, \frac{e^{i2\pi}}{\sqrt{r}\, e^{i\pi}(re^{i2\pi} + 1)}, \tag{13.54}$$

which can be rewritten as

$$\int_3 = \int_0^\infty dr\, \frac{1}{\sqrt{r}\,(r + 1)} = I. \tag{13.55}$$

Along path 4, we want to integrate around the branch point with a tiny circular contour whose radius will shrink to zero. If the radius is ϵ, along the contour $\underline{z} = \epsilon e^{i\theta}$, $d\underline{z} = i\epsilon e^{i\theta}d\theta$ and, because we are going clockwise this time, θ ranges from 2π to 0. This integral becomes

$$\int_4 = \lim_{\epsilon \to 0} \int_{2\pi}^0 d\theta\, \frac{i\epsilon e^{i\theta}}{\sqrt{\epsilon}\, e^{i\theta/2}(\epsilon e^{i\theta} + 1)}$$
$$= 0. \tag{13.56}$$

So the integration around the branch point vanishes. Remember this will not always be the case. There is no rule saying the integral around a branch point must vanish.

Combining the results of these four integrals with Equation 13.50 gives

$$\oint_{C_o} = 2\pi = \int_1 + \int_2 = 2I. \tag{13.57}$$

Therefore, our final result is

$$\int_0^\infty dx\, \frac{1}{\sqrt{x}\,(x + 1)} = \pi. \tag{13.58}$$

13.2 THE METHOD OF STEEPEST DESCENT

The method of steepest descent, also known as saddle point integration, is a technique for approximating line integrals of the form

$$I = \int_C d\underline{z}\, \underline{f}(\underline{z}), \tag{13.59}$$

where C is a path in the complex $\underline{z}$-plane, as shown in Figure 13.40. The method relies on analytic contour deformation to manipulate C so that only a small section on the contour actually contributes anything significant to the value of the integral. This deformation might be as shown in Figure 13.41. The modified contour C' passes
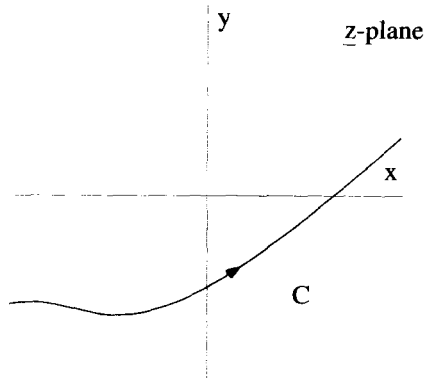
Figure 13.40 Initial Contour

through a point z_o, where $|f(z)|$ has a steep maximum which dwarfs the values of $f(z)$ at the other points on the contour. The path C' is called the path of steepest descent if no other contour can be drawn which has $|f(z)|$ decreasing more rapidly as one moves away from the maximum. Along this path, the integral can be approximated by the value obtained by integrating along C' for only a short distance near z_o,

$$I \approx \int_{z_o-\epsilon}^{z_o+\epsilon} dz\, f(z),$$

(13.60)

as indicated in Figure 13.42.
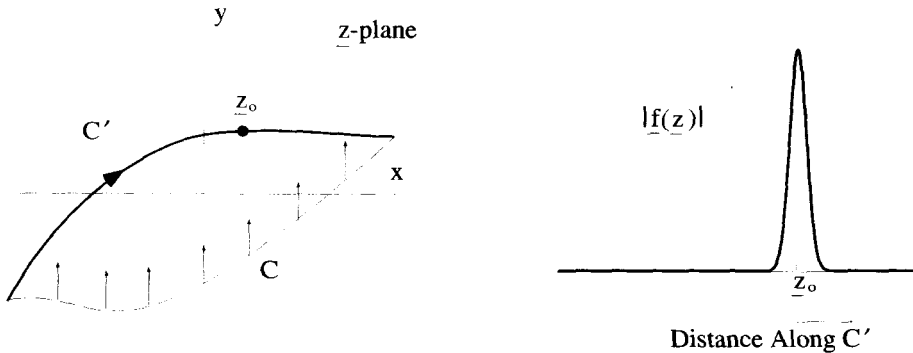
Figure 13.41 Steepest-Descent Contour
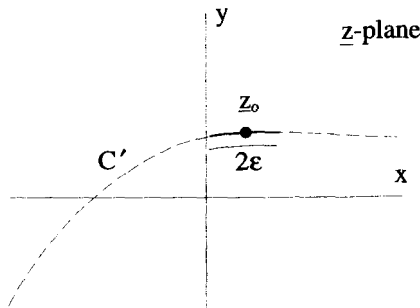
Figure 13.42   The Steepest-Descent Approximation

13.2.1   Finding the Local Maximum

To emphasize the steepness of the integrand and facilitate finding the local maximum, the integrand $\underline{f}(\underline{z})$ is placed in the form

$$\underline{f}(\underline{z}) = \underline{g}(\underline{z}) e^{\underline{w}(\underline{z})}, \qquad (13.61)$$

where $\underline{g}(\underline{z})$ is supposed to be a relatively slowly varying function of $\underline{z}$, while the exponential term $e^{\underline{w}(\underline{z})}$ changes magnitude rapidly. Given any $\underline{f}(\underline{z})$, the selection of $\underline{g}(\underline{z})$ and $\underline{w}(\underline{z})$ is not unique. You can always choose

$$\underline{g}(\underline{z}) = 1 \qquad (13.62)$$

and

$$\underline{w}(\underline{z}) = \ln[\underline{f}(\underline{z})]. \qquad (13.63)$$

In some cases, however, other choices are more appropriate.

Because $\underline{w}(\underline{z})$ is a complex function, we can write it as the sum of its real and imaginary parts:

$$\underline{w}(\underline{z}) = u(x, y) + iv(x, y), \qquad (13.64)$$

so that

$$|\underline{f}(\underline{z})| = |\underline{g}(\underline{z})| e^{u(x,y)}. \qquad (13.65)$$

Because $\underline{g}(\underline{z})$ is varying slowly, the magnitude of $\underline{f}(\underline{z})$ is controlled predominately by $u(x, y)$. Thus, a maximum in the magnitude of $\overline{f}(\underline{z})$ will occur at the maximum of $u(x, y)$, but will be much steeper due to the effect of the exponential function. Because $u(x, y)$ is a function of two variables, its extreme points occur when both

$$\frac{\partial u}{\partial x} = 0 \qquad (13.66)$$
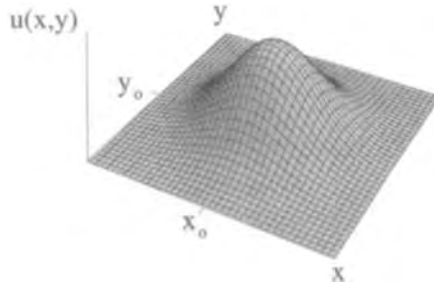
Figure 13.43 Local Maximum

and

$$\frac{\partial u}{\partial y} = 0. \tag{13.67}$$

If the values of x and y that allow the simultaneous solution of Equations 13.66 and 13.67 are x_o and y_o, the location in the complex $\underline{z}$-plane of an extremum of $u(x, y)$ is given by $\underline{z}_o = x_o + iy_o$.

To determine the nature of a particular extreme point $\underline{z}_o = x_o + iy_o$, the second derivatives of $u(x, y)$ must be investigated. If we define $u_{xx} \equiv \partial^2 u/\partial x^2$, $u_{yy} \equiv \partial^2 u/\partial y^2$, and $u_{xy} \equiv \partial^2 u/\partial x \partial y$, there are four general possibilities (see for example, Thomas and Finney, *Calculus and Analytical Geometry*):

$$u_{xx} < 0 \text{ and } u_{xx}u_{yy} - u_{xy}^2 > 0 \qquad \underline{z}_o \text{ is a maximum}$$

$$u_{xx} > 0 \text{ and } u_{xx}u_{yy} - u_{xy}^2 > 0 \qquad \underline{z}_o \text{ is a minimum}$$

$$u_{xx}u_{yy} - u_{xy}^2 < 0 \qquad \underline{z}_o \text{ is a saddle point}$$

$$u_{xx}u_{yy} - u_{xy}^2 = 0 \qquad \text{cannot be determined.} \tag{13.68}$$

You are no doubt familiar with the first two cases, which are simply local maxima and minima, as shown in Figures 13.43 and 13.44. The third case might be something new to you. For functions of two variables, it is possible to have an extremum which looks like a maximum in one direction, but a minimum from another. This type of
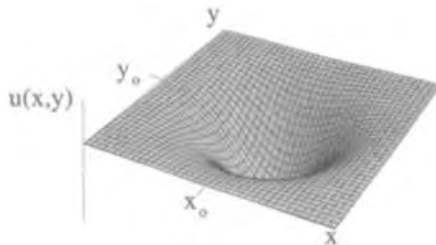
Figure 13.44 Local Minimum
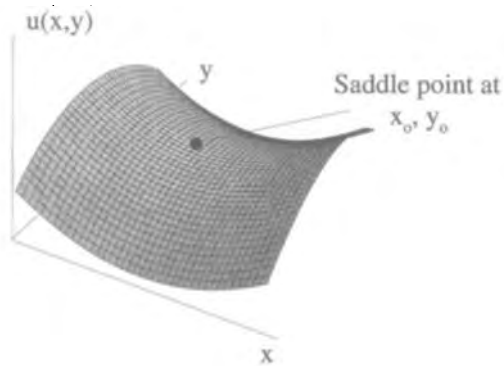
Figure 13.45   Extremum for Analytic Functions

extremum, is called a saddle point and is shown in Figure 13.45. The fourth condition listed in Equation 13.68 says that if the extreme point falls into this category, it is impossible to determine the nature of the extremum simply by looking at second derivatives. Higher-order terms must be considered.

Usually in contour integral problems, we are dealing with functions which are analytic everywhere, except at isolated points. If we impose the condition that $\underline{w}(\underline{z})$ be an analytic function, and consequently obeys the Cauchy-Riemann conditions, an interesting requirement is imposed on the extrema of $u(x, y)$. The Cauchy-Riemann conditions specify that

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \tag{13.69}$$

and

$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{13.70}$$

This means that

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \tag{13.71}$$

Therefore, if $u_{xx} > 0$, then $u_{yy} < 0$, and vice-versa. Looking back at the conditions of Equation 13.68, this means all the extrema for an analytic function fall into the third or fourth categories. Those in the third category are all saddle points. The ones in the fourth category occur when u_{xx} and u_{yy} are both zero. These also turn out to be saddle points of a more complicated type. We defer the discussion of these cases until the end of the chapter. When applying the method of steepest descent to a saddle point, the deformed contour needs to cross the saddle such that the contour goes through a maxima at $\underline{z}_o$, not a minima.

We already determined the extrema of $\underline{w}(\underline{z})$ are located where x and y satisfy Equations 13.66 and 13.67. If we are dealing with analytic functions, these equations also imply that $\partial v/\partial x$ and $\partial v/\partial y$ will also be zero at these points. Therefore, a simpler way to determine the location of the saddle points of analytic functions is by using the expression

$$\frac{du}{dx} + i\frac{dv}{dy} = \frac{d\underline{w}}{d\underline{z}} = 0 \qquad \text{at the saddle points.} \qquad (13.72)$$

13.2.2 Determining the Path of Steepest Descent

The path of steepest descent is the path through the saddle point that makes the peak of the curve, plotted in Figure 13.41, as narrow as possible. One simple way to visualize this is to think of the area around the saddle point as a mountain pass, and the path of steepest descent the shortest way through the pass. Alternatively, it is the path a skier would take to descend from the saddle point the fastest. Notice, the course the path takes is dependent on the details of the topology. It does not necessarily follow lines of constant y or constant x, and in general is not even a straight line.

A useful way to picture the path of steepest descent is on a two-dimensional contour plot, such as the one shown in Figure 13.46. Lines of constant $u(x, y)$ are plotted, with the solid lines used for $u(x, y) > u(x_o, y_o)$ and dashed lines for $u(x, y) < u(x_o, y_o)$. The saddle point (SP) is located at (x_o, y_o). The path of steepest descent passing through the saddle point has also been plotted. Notice how the path of steepest decent always intersects the dashed lines of constant $u(x, y)$ at right angles.

13.2.3 Evaluation of the Integral

Now we will attempt to approximate the contour integral along this path by only considering the contribution of a short section of the path around the saddle point.
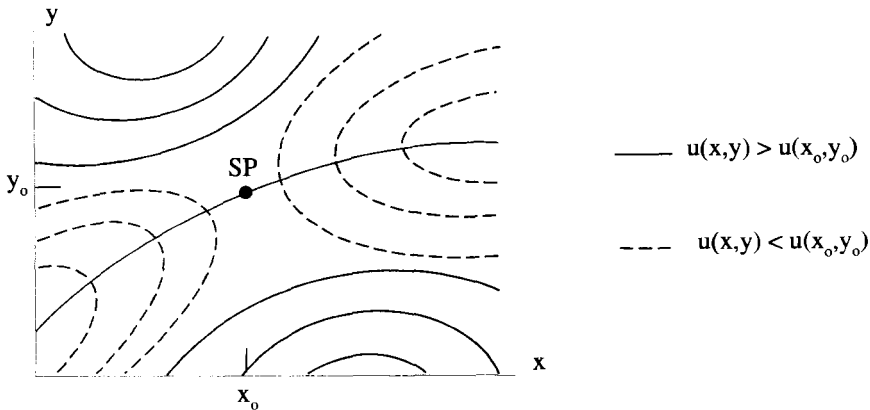
Figure 13.46 Lines of Constant $u(x, y)$
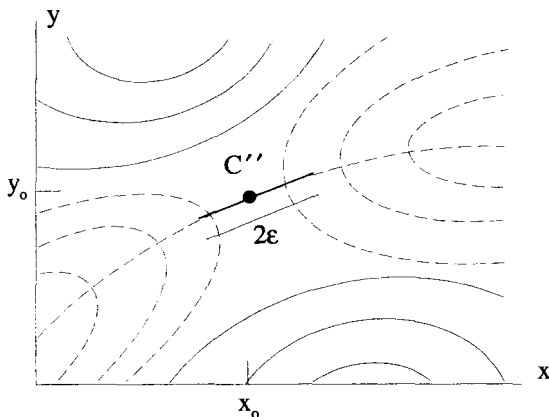
Figure 13.47   Contour for the Steepest-Descent Approximation

Near the saddle point, the path can be approximated by a straight-line segment, tangent to the actual path of steepest descent at the saddle point. This is shown as the path C'' in Figure 13.47. So the original integral, Equation 13.59 evaluated on the contour C, is exactly equivalent to the integral along C' of Figure 13.41, as long as the integrand is analytic between C and C'. This value is approximately equal to the integral along C'' of Figure 13.47:

$$I = \int_C d\underline{z}\, \underline{f}(\underline{z}) = \int_{C'} d\underline{z}\, \underline{f}(\underline{z}) \approx \int_{C''} d\underline{z}\, \underline{f}(\underline{z}). \qquad (13.73)$$

It is useful to modify the form of the integrand of Equation 13.61 slightly, by adding a constant factor $s > 0$ in front of the function $\underline{w}(\underline{z})$. This makes the integration we want to perform a function of s:

$$I(s) \approx \int_{C''} d\underline{z}\, \underline{g}(\underline{z}) e^{s\underline{w}(\underline{z})}. \qquad (13.74)$$

Essentially, the value of s is a measure of the validity of the approximation. As s increases, the steepness of the integrand's peak also increases, making the integral along C'' a better approximation to the integration along the original contour C.

Along the short contour C'', the value of $\underline{z}$ can be written

$$\underline{z} = \underline{z}_{sp} + \delta\, e^{i\alpha_o}, \qquad (13.75)$$

where $\underline{z}_{sp}$ is the saddle point position, δ is a small displacement, and α_o is a constant angle. If the length of C'' is 2ϵ, as shown in Figure 13.48, then δ ranges from $-\epsilon$ to $+\epsilon$ as $\underline{z}$ moves between the end points of C''. The value of α_o is the angle, with respect to the x-axis, that makes C'' tangent to the path of steepest descent. We will
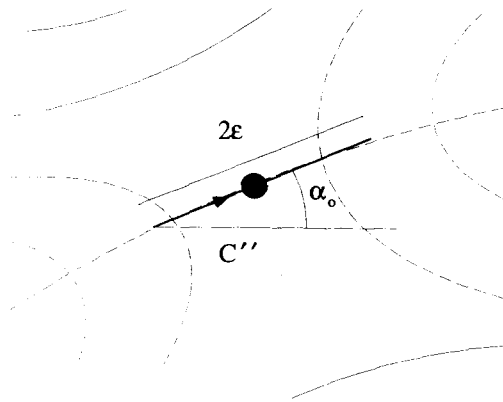
Figure 13.48 Definition of the Angle α_o

present a method for determining this angle shortly. We can now change the variable of integration of Equation 13.74 from $\underline{z}$ to δ by substituting Equation 13.75 and using

$$d\underline{z} = e^{i\alpha_o} d\delta. \tag{13.76}$$

The approximation in Equation 13.74 becomes

$$I(s) \approx e^{i\alpha_o} \int_{-\epsilon}^{+\epsilon} d\delta \, \underline{g}(\underline{z}_{sp} + \delta e^{i\alpha_o}) e^{s\underline{w}(\underline{z}_{sp} + \delta e^{i\alpha_o})}. \tag{13.77}$$

Since we assumed $\underline{g}(\underline{z})$ to be a slowly varying function, it can be approximated by a constant value $\underline{g}(\underline{z}_{sp})$ for the C'' integration and therefore can be pulled outside of the integral. When a Taylor expansion for $\underline{w}(\underline{z})$ is made about $\underline{z}_{sp}$, we obtain the series

$$\underline{w}(\underline{z}) = \underline{w}(\underline{z}_{sp}) + \left.\frac{d\underline{w}}{d\underline{z}}\right|_{\underline{z}_{sp}} (\underline{z} - \underline{z}_{sp}) + \frac{1}{2} \left.\frac{d^2\underline{w}}{d\underline{z}^2}\right|_{\underline{z}_{sp}} (\underline{z} - \underline{z}_{sp})^2 + \cdots, \tag{13.78}$$

which, written in terms of δ, becomes

$$\underline{w}(\underline{z}_{sp} + \delta e^{i\alpha_o}) = \underline{w}(\underline{z}_{sp}) + \frac{1}{2} \left.\frac{d^2\underline{w}}{d\underline{z}^2}\right|_{\underline{z}_{sp}} (\delta e^{i\alpha_o})^2 + \cdots. \tag{13.79}$$

Notice because $d\underline{w}/d\underline{z}|_{\underline{z}_{sp}} = 0$, the lowest-order δ-dependence is δ^2. If we approximate the function $\underline{w}(\delta)$ by taking only terms up to the lowest order in δ, and combine this with the constant approximation for $g(\underline{z})$, we obtain

$$I(s) \approx \underline{g}(\underline{z}_{sp}) e^{i\alpha_o} e^{s\underline{w}(\underline{z}_{sp})} \int_{-\epsilon}^{\epsilon} d\delta \, e^{\frac{s}{2} \frac{d^2\underline{w}}{d\underline{z}^2}\big|_{\underline{z}_{sp}} \delta^2 e^{i2\alpha_o}}. \tag{13.80}$$

This result assumes that the second derivative, which appears in the exponent of the integral, is not zero. Later in the chapter, we will briefly discuss a situation where this is not the case.

The only task which remains is to determine the angle α_o. The path of steepest descent follows the line which is perpendicular to the line of constant $u(x, y)$ that passes through the saddle point. From our previous work with analytic functions of complex variables, we know that this path is also along a line of constant $v(x, y)$, because the lines of constant $v(x, y)$ are always perpendicular to lines of constant $u(x, y)$. This implies that, along the path of steepest descent, the imaginary part of $\underline{w}(\underline{z})$ is a constant equal to its value at the saddle point. Consequently, on the path of steepest descent, where we have approximated $\underline{w}(\underline{z})$ by

$$\underline{w}(\underline{z}_{sp}) + \frac{1}{2} \frac{d^2 \underline{w}}{d\underline{z}^2}\bigg|_{\underline{z}_{sp}} (\delta e^{i\alpha_o})^2, \tag{13.81}$$

the imaginary part of $\underline{w}(\underline{z})$ is entirely contained in the $\underline{w}(\underline{z}_{sp})$ term and

$$\frac{1}{2} \frac{d^2 \underline{w}}{d\underline{z}^2}\bigg|_{\underline{z}_{sp}} (\delta e^{i\alpha_o})^2 \tag{13.82}$$

must be a pure real quantity. Since δ is a pure real quantity, that means that

$$\frac{d^2 \underline{w}}{d\underline{z}^2}\bigg|_{\underline{z}_{sp}} e^{i2\alpha_o} \tag{13.83}$$

must be a pure real quantity. In general, $d^2\underline{w}/d\underline{z}^2|_{\underline{z}_{sp}}$ is a complex quantity, which we write using the exponential representation as

$$\frac{d^2 \underline{w}}{d\underline{z}^2}\bigg|_{\underline{z}_{sp}} = W_o'' e^{i\theta_o}. \tag{13.84}$$

In order for Equation 13.83 to be pure real,

$$\theta_o + 2\alpha_o = \pm n\pi \qquad n = 0, 1, 2, \ldots. \tag{13.85}$$

If n is an odd integer, Equation 13.80 becomes

$$I(s) \approx \underline{g}(\underline{z}_{sp}) e^{i\alpha_o} e^{s\underline{w}(\underline{z}_{sp})} \int_{-\epsilon}^{\epsilon} d\delta\, e^{-\frac{s}{2} W_o'' \delta^2}, \tag{13.86}$$

while if n is zero or an even integer,

$$I(s) \approx \underline{g}(\underline{z}_{sp}) e^{i\alpha_o} e^{s\underline{w}(\underline{z}_{sp})} \int_{-\epsilon}^{\epsilon} d\delta\, e^{+\frac{s}{2} W_o'' \delta^2}. \tag{13.87}$$

Because we required s to be a positive real quantity, the integrand of Equation 13.86 drops off as δ moves from zero, that is, as $\underline{z}$ moves away from the saddle point. Therefore, this equation is an approximatation for the integration along the path of steepest descent. In contrast, the integrand of Equation 13.87 grows as δ moves from zero, and approximates the integration along a path of steepest ascent. This is not the path we want. With $n = 1$, α_o must be given by

$$\alpha_o = \frac{-\theta_o \pm \pi}{2}.$$
(13.88)

The other odd values of n in Equation 13.85 will not generate any new values for α_o. The ambiguity of the $\pm$ sign must be determined by the direction taken by the undeformed contour C and the topology of $u(x, y)$, which controls the way C is deformed to C'. The deformation must be such that C' stays in the "lowland," except when it crosses the saddle point.

Equation 13.86 is almost the final result. The effect of the s parameter is now clear. The integrand of this expression is a Gaussian whose width is controlled by s. The larger s becomes, the narrower the Gaussian spike becomes, making the approximation more valid. This effect allows one last simplification of the result. If the Gaussian is narrow enough, the contributions from extending the limits of integration to minus and plus infinity will be negligible. With this approximation, the integration in Equation 13.86 can be evaluated to give the result

$$I(s) \approx \left[\underline{g}(\underline{z}_{sp}) e^{s\underline{w}(\underline{z}_{sp})} e^{i\alpha_o} \right] \sqrt{\frac{2\pi}{sW_o''}}.$$
(13.89)

After all these approximations, you might wonder if the result could possibly bear any resemblance to the actual value of the original integral. If s is large enough and the deformation to a path of steepest descent can be accomplished, the approximations turn out to produce, in many cases, amazingly accurate results. This will be demonstrated in the examples that follow and in the exercises at the end of this chapter.

Example 13.9   One frequent application of the method of steepest descent is the evaluation of Laplace transforms and their inverses. For example, the transform

$$\underline{F}(\underline{s}) = \int_0^\infty dt\, f(t) e^{-\underline{s}t}$$
(13.90)

can be approximated using the method of steepest descent for positive real values of $\underline{s}$. The first step is to place the integral in steepest descent form, i.e.,

$$\int_C d\underline{z}\, \underline{g}(\underline{z}) e^{s\underline{w}(\underline{z})}.$$
(13.91)

To accomplish this, convert the transformation to an integral in the complex plane by letting t become $\underline{z}$. The initial integration is then changed into a contour integral
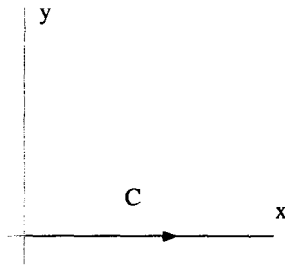
Figure 13.49   Contour for Laplace Transforms

along the positive real z axis,

$$\underline{F}(s) = \int_C d\underline{z} f(\underline{z}) e^{-s\underline{z}},$$ (13.92)

as shown in Figure 13.49.

Now we need to identify the two functions $g(\underline{z})$ and $\underline{w}(\underline{z})$. There are several possibilities. Probably the simplest approach is to set $g(\underline{z}) = 1$ and

$$f(\underline{z}) e^{-s\underline{z}} = e^{s\underline{w}(\underline{z})}$$ (13.93)

so that

$$\underline{w}(\underline{z}) = \frac{\ln[f(\underline{z})] - s\underline{z}}{s}.$$ (13.94)

A more general approach would allow $f(\underline{z})$ to be broken up into two parts, a slow-varying part $g(\underline{z})$ and a fast-varying part $h(\underline{z})$. Then we would set

$$f(\underline{z}) = g(\underline{z}) h(\underline{z}),$$ (13.95)

with

$$\underline{w}(\underline{z}) = \frac{\ln[h(\underline{z})] - s\underline{z}}{s}.$$ (13.96)

Sticking with our original choice of Equation 13.94, the Laplace transform integral becomes

$$\underline{F}(s) = \int_C d\underline{z}\, e^{s\{\frac{\ln[f(\underline{z})]}{s} - \underline{z}\}}.$$ (13.97)

The next step is to find the saddle point. This is done be setting $d\underline{w}/d\underline{z} = 0$ and solving for $\underline{z}$. In this case, the saddle point is determined from the condition

$$\frac{df(\underline{z})}{d\underline{z}} = s\, f(\underline{z}).$$ (13.98)

To complete the approximation, we need to determine the second derivative of $\underline{w}(\underline{z})$ at the saddle point, and identify the values of W_o'' and α_o to use in Equation 13.89. These steps are straightforward, but the details are of course determined by the nature of the function $\underline{f}(\underline{z})$.

Example 13.10 As a more specific example of the use of the method of steepest descent, we will work out an approximate expression for the Γ-function, which is defined by the integral expression

$$\Gamma(s + 1) = \int_0^\infty d\rho\, \rho^s e^{-\rho}. \tag{13.99}$$

For positive integer values of s, this turns out to be intimately connected with the factorial function:

$$\Gamma(s + 1) = s!. \tag{13.100}$$

The Γ-function can be used to define factorials of negative and noninteger numbers.

To apply the method of steepest descent, we extend the integral into the complex plane by replacing ρ with $\underline{z}$,

$$\Gamma(s + 1) = \int_C d\underline{z}\, \underline{z}^s e^{-\underline{z}}, \tag{13.101}$$

where again C is along the positive real axis, the same contour shown in Figure 13.49.

The integral for the Γ-function must now be put into steepest descent form, so that $\underline{w}(\underline{z})$ and $\underline{g}(\underline{z})$ can be identified. Remember the desired form is

$$\Gamma(s + 1) = \int_C d\underline{z}\, \underline{g}(\underline{z}) e^{s\underline{w}(\underline{z})}. \tag{13.102}$$

Neither $\underline{z}^s$ nor $e^{-\underline{z}}$ appear to be slowly varying functions and so we will set $\underline{g}(\underline{z}) = 1$. This is always a safe choice. That means that $\underline{w}(\underline{z})$ in Equation 13.102 becomes

$$\underline{w}(\underline{z}) = \ln(\underline{z}) - \frac{\underline{z}}{s}. \tag{13.103}$$

Taking the derivative of this function with respect to $\underline{z}$ and setting it equal to zero gives the location of the saddle point:

$$\underline{z}_{sp} = s. \tag{13.104}$$

Because s is supposed to be a positive real quantity, the saddle point is on the original contour, and no deformation is necessary to make the contour cross it.

It is pretty clear that α_o should be zero, but let's see how this falls out of the method. To find α_o we first evaluate the second derivative of $\underline{w}(\underline{z})$ at the saddle point.

This gives

$$\frac{d^2\underline{w}}{d\underline{z}^2}\bigg|_{\underline{z}_{sp}} = -\frac{1}{\underline{z}_{sp}^2} = -\frac{1}{s^2} = \frac{1}{s^2}e^{i\pi}. \qquad (13.105)$$

This means that

$$W_o'' = \frac{1}{s^2} \qquad (13.106)$$

and

$$\theta_o = \pi. \qquad (13.107)$$

With this value for θ_o, Equation 13.88 gives two values for α_o, $\alpha_o = 0$ and $\alpha_o = -\pi$. If we were to use the value $\alpha_o = -\pi$, the contour would pass through the saddle point from right to left. Since the original integration was from left to right, the $\alpha = 0$ solution should be selected.

With these values for $\underline{z}_{sp}$, W_o'', α_o, and the $\underline{w}(\underline{z})$ and $\underline{g}(\underline{z})$ functions, Equation 13.89 gives the result:

$$\Gamma(s+1) \approx \sqrt{2s\pi}\, s^s e^{-s}. \qquad (13.108)$$

This is the simplest form of Stirling's approximation.

13.2.4 Higher-Order Saddles

Back in Equation 13.80, where we performed a Taylor series expansion of the function $\underline{w}(\underline{z})$ around the saddle point, we assumed the first nonzero term involved the second derivative. The result of Equation 13.89 and the validity of all of our examples relied
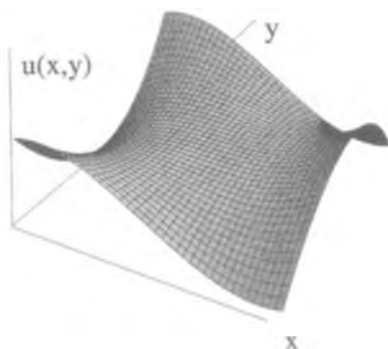
Figure 13.50   A Three-Legged Saddle

on this fact. An interesting situation occurs if the second derivative, evaluated at the saddle point, turns out to be zero. Then we must go to higher-order terms in the Taylor series to get the lowest δ dependent term needed for Equation 13.80. Interesting saddle structures occur in these cases. They are saddles for riders with more than two legs! An example of the saddle formed when the first nonzero term is the third derivative is shown in Figure 13.50. Notice how a single, straight-line contour cannot approximate the line of steepest descent near the saddle point in this case.

EXERCISES FOR CHAPTER 13

1. Consider the complex function

$$\underline{w} = (\underline{z}^2 - 1)^{1/3}.$$

   (a) How many $\underline{z}$-plane Riemann sheets are required to describe this function?
   (b) Locate the branch points and branch cuts in the $\underline{z}$-plane and show how the $\underline{z}$-plane Riemann sheets are connected.
   (c) How many $\underline{w}$-plane Riemann sheets are required to describe this function?
   (d) Locate the branch points and branch cuts in the $\underline{w}$-plane and show how the $\underline{w}$-plane Riemann sheets are connected.

2. The function

$$\underline{w} = (\underline{z}^2 - 9)^{1/2}$$

is represented on two sheets in the $\underline{z}$-plane and two sheets in the $\underline{w}$-plane. Place the cut on the real axis of the $\underline{z}$-plane and indicate where each sheet of the $\underline{z}$-plane maps onto the two $\underline{w}$-plane sheets.

3. Consider the complex function

$$\underline{w} = (\underline{z} - 1)^{1/2}(\underline{z} + 1)^{1/3}.$$

   (a) Locate the branch points and branch cuts of this function in the $\underline{z}$-plane.
   (b) How many Riemann sheets are required for the $\underline{z}$-plane?
   (c) How are these Riemann sheets connected?
   (d) How many Riemann sheets are required for the $\underline{w}$-plane?

4. Consider the complex function

$$\underline{w} = \ln\left(\frac{\underline{z} + 1}{\underline{z} - 1}\right).$$

   (a) Identify the branch points and locate the branch cut(s) in the $\underline{z}$-plane for this function.

(b) How many z-plane Riemann sheets are required to represent this function?

(c) How are the Riemann sheets connected?

(d) Draw several closed contours, which span multiple sheets in the z-plane.

5. Locate the branch points and branch cuts in the z-plane and describe the connection of the Riemann sheets for the following functions:

   i. $w = \ln(z^2 - 1)$.

   ii. $w = z \ln z$.

   iii. $w = (z^2 - z)^{1/2}$.

   iv. $w = z^{1/2} + (z - 1)^{1/2}$.

6. For each of the following functions: (1) Identify the number of Riemann sheets in the complex z-plane and the complex w-plane. (2) Locate all the branch points and position the branch cuts. (3) Show where the z-plane maps onto the w-plane and vice versa. (4) Indicate how the Riemann sheets are connected:

   i. $w = z^{1/4}$.

   ii. $w = z^{2/3}$.

   iii. $w = \cos z$.

   iv. $w = (\ln z)^2$.

   v. $w = \ln(z^2 + 1)$.

7. The function $(z - 1)^{1/2}$ has branch points at $z = 1$ and at infinity. For this problem, take the cut to be along the positive real axis and evaluate the integral

$$\int_C dz\,(z - 1)^{1/2}.$$

The contour C is the circular path of radius 2, centered at $z = 0$, that starts just above the cut and ends just below the cut on the first Riemann sheet of the z-plane, as shown below:
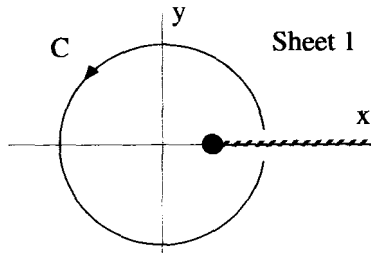
To evaluate this integral show that the contour can be deformed to the one shown below:
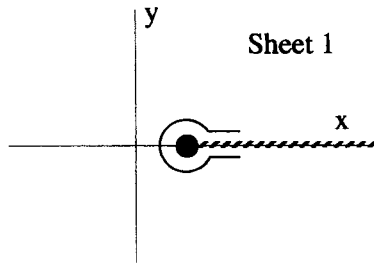
Sheet 1

Repeat the evaluation of the integral if C is assumed to be on the second Riemann sheet of the $\underline{z}$-plane.

8. Consider the real integral

$$I = \int_0^\infty dx \, \frac{\ln x}{x^2 + 1}.$$

Indicate a closed contour in the complex $\underline{z}$-plane that can be used to evaluate this integral. Explain your reasoning by showing the pole(s), branch points, and branch cut(s) of the integrand and discussing the contributions from the segments used to accomplish closure. You do not have to actually evaluate the integral.

9. Evaluate

$$\int_0^\infty dx \, \frac{\sqrt{x}}{(x + 1)^2}$$

by extending the integral into the complex plane and using the contour C shown below.
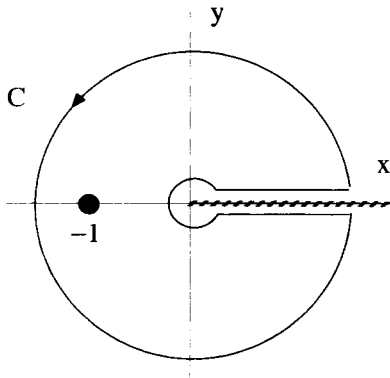
10. Consider the complex function

$$\underline{w} = \frac{1}{(\underline{z}^2 - \underline{z})^{1/2}(\underline{z} + 2)}.$$

(a) Identify the branch cuts, branch points Riemann sheets, and singularities of this function in the $\underline{z}$-plane.

(b) Now consider the real integral

$$\int_0^1 \frac{dx}{(x^2 - x)^{1/2}(x + 2)}.$$

Extend this integral into the complex plane and identify an appropriate contour to evaluate it.

(c) Evaluate the integral in part (b).

11. Describe the branch cuts and branch points of the function $(1 - \underline{z}^2)^{1/2}$ and evaluate the real integral

$$\int_{-1}^1 \frac{dx}{(x - 2)\sqrt{1 - x^2}}.$$

12. Evaluate the real integral

$$\int_{-1}^1 \frac{dx}{(x^2 - 1)^{1/2}(ax + b)} \qquad b > a > 0.$$

13. Consider the function $\underline{w} = \underline{z}^2 + 1$.

(a) Determine the location of the saddle point for this function.

(b) Expand $\underline{w}(\underline{z})$ in a Taylor series about the saddle point and determine α_o, the angle of a straight line passing through the saddle point tangent to the path of steepest descent.

(c) Repeat steps (a) and (b) above for the function $\underline{w} = \underline{z}^3 + 1$.

14. Consider the function $\underline{z}$

$$\underline{w} = \frac{\underline{z}^2}{2} + \cos \underline{z}.$$

(a) Show that this function satisfies the Cauchy-Riemann conditions.

(b) Identify the location of a saddle point.

(c) Describe the paths of steepest ascent and descent through this saddle point.

15. Consider the function

$$\underline{w} = \underline{z}^3 + 3\underline{z}.$$

(a) Where in the complex $\underline{z}$-plane does this function have saddle points?

(b) Identify the lines in the $\underline{z}$-plane along which Real $[\underline{w}] = 0$.

(c) The lines determined in part (b) divide the $\underline{z}$-plane into six regions. For each region, indicate whether the real part of $\underline{w}$ is greater or less than zero.

(d) Near each of the saddle points, determine the directions of the paths of steepest descent and steepest ascent through these saddle points.

(e) Using the method of steepest descent, determine the approximate value of the real integral

$$\int_0^\infty dx\, e^{-s(x^3+3x)},$$

where s is a large, positive real number.

16. Let

$$\underline{w} = \sin \underline{z}.$$

(a) Where are the saddle points of this function in the $\underline{z}$-plane?

(b) Identify the paths of steepest descent and ascent through these saddle points.

(c) Make a plot of the lines of constant Real $[\underline{w}]$ in the complex plane and label several of the contours to clearly show the topology. Also draw the lines of steepest descent and ascent on this plot.

17. Consider the function

$$\underline{w} = \sin \underline{z} - \underline{z}.$$

(a) Show that this function has saddle points at $\underline{z} = 2\pi n$ where $n = 0, \pm 1, \pm 2, \dots$.

(b) Expand the function in a Taylor series about the saddle point at $\underline{z} = 0$. What is the second derivative of $\underline{w}$ at this point?

(c) What is the value of the function at $\underline{z} = 0$? Use this value to divide the $\underline{z}$-plane into regions around $\underline{z} = 0$, where the real part of the function is either larger than or smaller than its real part at $\underline{z} = 0$.

(d) Determine all the possible paths of steepest descent and steepest ascent that pass through $\underline{z} = 0$.

18. Use the method of steepest descent to approximate the value of the integral

$$I(x) = \int_0^\infty dt\, e^{xt - e^t}$$

for large, positive x.

19. Consider the real integral

$$\int_0^\infty dx\, xe^{-sx^2},$$

where s is a real positive number.

(a) Evaluate this integral exactly using the fact that the integrand is an exact differential.

(b) Now evaluate the integral by the method of steepest descent and compare this approximate answer to the exact expression obtained in part (a).

20. Use the method of steepest descent to approximate the Laplace transform of

$$f(t) = \begin{cases} t & t > 0 \\ 0 & t < 0 \end{cases}$$

and compare it to the exact transform. Now repeat the same process for the transform of

$$f(t) = \begin{cases} t^n & t > 0 \\ 0 & t < 0 \end{cases},$$

where n is any positive integer.

21. Consider the integral

$$I(x) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} d\underline{s} \, \frac{e^{x(\underline{s} - \underline{s}^{1/2})}}{\underline{s}},$$

where the path of integration is along a vertical line to the right of the origin in the $\underline{s}$-plane.

(a) Locate the branch points of the integrand in the $\underline{s}$-plane.

(b) Place the branch cut so that the path of integration will not cross it.

(c) Evaluate $I(x)$ for negative x.

(d) For large positive x, use the method of steepest descent to approximate $I(x)$.

22. According to transform tables, the inverse Laplace transform of

$$\frac{1}{\underline{s}^{3/2}} e^{-\alpha/\underline{s}}$$

is 0 for $t < 0$, and equals

$$\frac{1}{\sqrt{\pi\alpha}} \sin\left(2\sqrt{\alpha t}\right)$$

for $t > 0$. Use the method of steepest descent to approximate

$$\frac{1}{2\pi i} \int_{\mathcal{L}} d\underline{s} \, \frac{e^{-\alpha/\underline{s}} e^{\underline{s}t}}{\underline{s}^{3/2}}.$$

Contour plots for the constant real part of $\underline{w}(\underline{s})$ help to see how to deform the path of integration. When you deform the Laplace contour $\mathcal{L}$, remember to take into account the branch cut, branch points, and the singularity of the integrand. The integrand has a slowly varying function of $\underline{s}$, and both saddle points of

$w(\underline{s})$ should be used. How does your approximate answer compare to the correct answer for large values of t? What would have happened if you had used only one saddle point? If possible, use a computer to make a 3D plot of the real part of $w(\underline{s})$.

23. Use the method of steepest descent to find an approximation for the inverse Laplace transform of

$$\underline{F}(\underline{s}) = \frac{1}{\underline{s}^{1/2}}.$$

Show the paths of steepest ascent and descent through the saddle point(s) and discuss the deformation of the path of integration. Also, include comments on any branch points and branch cuts in your solution. Compare your answer to the exact value of the inversion, which is

$$f(t) = \begin{cases} 0 & t < 0 \\ 1/\sqrt{\pi t} & t > 0 \end{cases}.$$

Now repeat this process for the inversion of

$$\underline{F}(\underline{s}) = \frac{1}{\underline{s}^{1/2}(\underline{s} - 1)}.$$

14

TENSORS IN NON-ORTHOGONAL COORDINATE SYSTEMS

This chapter continues our discussion of tensors. It begins with a brief review of tensor transformation properties, which was covered in more detail in Chapter 4. Then we tackle the complications that arise in non-orthonormal coordinate systems. The ideas of contravariance, covariance, and the metric tensor are developed to handle inner products in non-orthonormal systems.

14.1 A BRIEF REVIEW OF TENSOR TRANSFORMATIONS

In Chapter 4, we discussed how a tensor is defined by its behavior under coordinate transformations. With only a hint of sarcasm, the definition was given by the statement, "A tensor is any quantity that transforms as a tensor." What this means is that transformation rules are enough to give tensors all of their special properties. If an object transforms between coordinate systems using tensor transformation rules, you can legitimately say the object is a tensor.

Remember, the elements of a tensor can be transformed using a transformation matrix whose elements can be obtained from the equations relating the coordinates of the two systems under consideration. For transformations between Cartesian systems, the elements of this transformation matrix $[a]$ are given by

$$a_{ij} = (\hat{\mathbf{e}}_i' \cdot \hat{\mathbf{e}}_j) = \frac{\partial x_i'}{\partial x_j}. \tag{14.1}$$

In this equation, the original system has coordinates x_i and basis vectors $\hat{\mathbf{e}}_i$. The system being transformed to, the primed system, has coordinates x_i' and basis vectors $\hat{\mathbf{e}}_i'$. For orthonormal coordinate systems, the inverse of this transformation matrix is always

given by its transpose

$$a_{ij}^{-1} = a_{ji}. \tag{14.2}$$

An arbitrary tensor of rank n can be expressed in either the unprimed or primed coordinate system as

$$\overline{\overline{\mathbf{T}}} = T_{ijk...}\hat{\mathbf{e}}_i\hat{\mathbf{e}}_j\hat{\mathbf{e}}_k \ldots = T'_{rst...}\hat{\mathbf{e}}'_r\hat{\mathbf{e}}'_s\hat{\mathbf{e}}'_t \ldots, \tag{14.3}$$

where there are n subscripts and n basis vectors in each term. $T_{ijk...}$ and $T'_{rst...}$ are the elements of the tensor in the unprimed and primed coordinate systems, respectively. The two sets of elements are related to each other via the transformation matrix equation

$$T'_{rst...} = T_{ijk...}a_{ri}a_{sj}a_{tk} \ldots, \tag{14.4}$$

where the matrix $[a]$ appears n times. The inverse transformation is

$$T_{rst...} = T'_{ijk...}a_{ir}a_{js}a_{kt} \ldots. \tag{14.5}$$

Our fundamental assertion is that any quantity that transforms in the manner described by Equation 14.4 is by definition a tensor.

Because a vector is a first rank tensor, vector component transformations are also described by Equations 14.4 and 14.5. If we write the vector in the two coordinate systems as

$$\overline{\mathbf{V}} = V_i\hat{\mathbf{e}}_i = V'_r\hat{\mathbf{e}}'_r, \tag{14.6}$$

the relation between the components is given by

$$V'_r = V_i a_{ri} \tag{14.7}$$

and the inverse

$$V_r = V'_i a_{ir}. \tag{14.8}$$

Scalars are invariant under a coordinate transformation. We can think of a scalar as a zero rank tensor. The single element of a zero rank tensor has no subscripts and is not combined with any basis vectors. Equation 14.4 reduces to

$$S = S' \tag{14.9}$$

where S (or S') is the single element of the scalar.

Example 14.1 As an example of using transform properties to identify an object as a tensor, consider the Kronecker-δ. Recall that this symbol was introduced to form dot products in orthonormal coordinate systems. The dot product between two vectors, $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$, written in two different orthonormal coordinate systems, one unprimed and

one primed, can be written

$$\overline{A} \cdot \overline{B} = A_i B_j \delta_{ij} = A'_r B'_s \delta'_{rs}. \qquad (14.10)$$

Now we already know that δ_{ij} and δ'_{rs} can both be expressed as unit matrices, that is as [1]. For the purposes of this discussion, however, let's look at the consequences of just requiring that both expressions for the inner product in Equation 14.10 be equal, and that A_i and B_i are vector components, and therefore transform according to Equation 14.7. Substituting into Equation 14.10 for A'_r and B'_s gives

$$A_i B_j \delta_{ij} = a_{ri} A_i a_{sj} B_j \delta'_{rs}.$$
$$= A_i B_j a_{ri} a_{sj} \delta'_{rs}. \qquad (14.11)$$

Because this expression must be true for any $\overline{A}$ and $\overline{B}$, we can write

$$\delta_{ij} = a_{ri} a_{sj} \delta'_{rs}. \qquad (14.12)$$

Inverting this expression gives

$$\delta'_{ij} = a_{ir} a_{js} \delta_{rs}. \qquad (14.13)$$

Comparing Equations 14.12 and 14.13 with Equations 14.4 and 14.5, it can be seen that the elements of the Kronecker-δ transform like the elements of a second-rank tensor. Therefore, the Kronecker-δ symbol *is* a second-rank tensor, which can be expressed with basis vectors as

$$\overline{\overline{\delta}} = \delta_{ij} \hat{e}_i \hat{e}_j = \delta'_{ij} \hat{e}'_i \hat{e}'_j. \qquad (14.14)$$

14.2 NON-ORTHONORMAL COORDINATE SYSTEMS

Up to this point, we have dealt only with orthonormal coordinate systems. In Cartesian systems, the basis vectors $\hat{e}_i$ are independent of position and orthonormal, so $\hat{e}_i \cdot \hat{e}_j = \delta_{ij}$. In curvilinear systems, the basis vectors $\hat{q}_i$ are no longer independent of position, but they are still orthonormal, so $\hat{q}_i \cdot \hat{q}_j = \delta_{ij}$. Now we will consider non-orthonormal systems. To distinguish these systems, we will label the basis vectors of non-orthonormal coordinate systems as $\hat{g}_i$, and the non-orthonormality condition becomes $\hat{g}_i \cdot \hat{g}_j \neq \delta_{ij}$. To keep the discussion and derivations as simple as possible in this chapter, we will limit ourselves to coordinate systems which have basis vectors that do not vary with position. This is obviously not the most general type of non-orthonormal coordinate system, but it is enough to demonstrate the important ideas of contravariance, covariance, and the metric tensor. In Appendix F, we present some of the generalizations required to describe non-orthonormal systems with position-dependent basis vectors.

In physics, non-orthonormal coordinate systems appear, for example, in both special and general relativity. The basic postulate of special relativity is that the speed of light c_o is the same in all inertial reference frames. As a consequence of
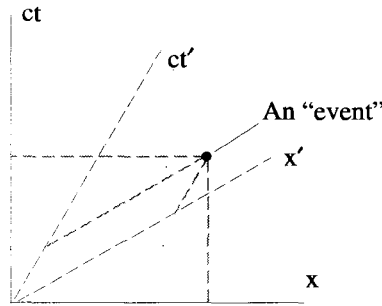
Figure 14.1 The Coordinate Systems of Special Relativity

this postulate, both the position and time coordinates of some physical phenomenon (an "event") change as you change reference frames. This is very similar to how the components of a vector change as you transform the coordinate axes used to describe them. If we restrain the motion to be along a single spatial direction, an event can be described by two coordinates, a single position coordinate and time. As will be shown, the observation of an event in two different coordinate systems, one primed and one unprimed, can be plotted as a single point using the combined set of axes as shown in Figure 14.1. By taking its components with respect to both coordinate axes, we can get all the relationships imposed by special relativity. Notice how the x and ct axes intersect at right angles, but the x' and ct' axes do not. While the unprimed system appears to be orthogonal, the primed system looks like a non-orthogonal, skewed system. We will discuss this picture, and its interpretation for special relativity, in some detail in the sections that follow.

The basic postulate of general relativity is that gravity and acceleration are equivalent. Events observed in a gravitational field appear as if they were being observed in an accelerating coordinate system. This implies that light propagating through the gravitational field of a massive object, such as a star, should be bent, as depicted in Figure 14.2. This would cause the apparent position of a star to deviate from its actual
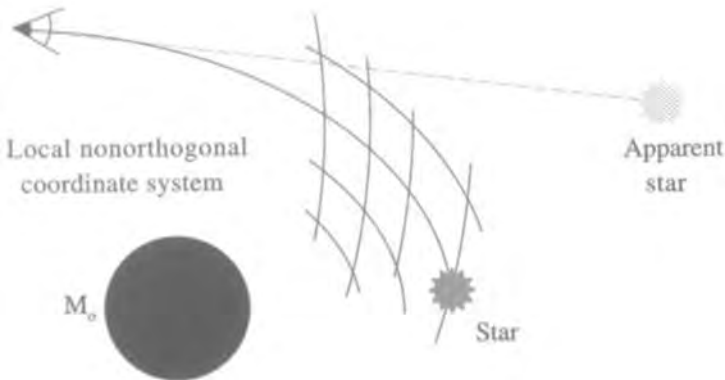
Figure 14.2 A Coordinate System of General Relativity

location. This phenomenon was first observed by Arthur Eddington, who measured the small deflection of stars by the sun during the total eclipse of 1919. The paths that light rays follow through space are called geodesics. A natural choice for the grid lines of a localized, coordinate system follows these geodesics, as shown in Figure 14.2. The basis vectors in such a coordinate system are both non-orthonormal and spatially dependent. We will not discuss examples of this type of system, but instead restrict our comments to "skewed" systems, where the basis vectors are not orthonormal, but are spatially invariant. The interested reader can investigate the very heavy book *Gravitation* by Misner, Thorne, and Wheeler for an introduction to general relativity.

14.2.1   A Skewed Coordinate System

Consider the two-dimensional, Cartesian system (x_1, x_2) and the non-orthonormal, primed system (x_1', x_2') shown in Figure 14.3. Two pairs of basis vectors and an arbitrary vector $\overline{V}$ are also represented in the figure. The basis vectors of the unprimed system are orthonormal:

$$\hat{e}_i \cdot \hat{e}_j = \delta_{ij}. \tag{14.15}$$

The basis vectors of the skewed system are taken to be unit vectors,

$$\hat{g}_1' \cdot \hat{g}_1' = \hat{g}_2' \cdot \hat{g}_2' = 1, \tag{14.16}$$

but they are not orthonormal:

$$\hat{g}_i' \cdot \hat{g}_j' \neq \delta_{ij}. \tag{14.17}$$
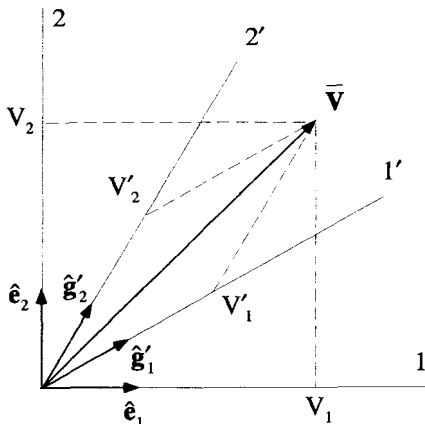
Figure 14.3   An Orthonormal and a Skewed Coordinate System

As you will see, the formalism that is developed here does allow for systems with basis vectors that are not unit vectors. However, to start out as simply as possible, we assume the basis vectors of the skewed system satisfy Equation 14.16.

In the orthonormal system, the vector $\overline{V}$ can be expressed as the sum of its axis parallel projected components, shown in Figure 14.3, coupled with the corresponding basis vectors:

$$\overline{V} = V_1\,\hat{e}_1 + V_2\,\hat{e}_2. \tag{14.18}$$

These vector components are just the projected lengths of $\overline{V}$ along the axes of the unprimed system and can be determined with simple trigonometry or by following straightforward vector manipulation. A particular component is obtained by the dot product between the vector $\overline{V}$ and the corresponding basis vector. For example, to find V_1:

$$\begin{aligned}
\overline{V} \cdot \hat{e}_1 &= (V_1\,\hat{e}_1 + V_2\,\hat{e}_2) \cdot \hat{e}_1 \\
&= V_1(\hat{e}_1 \cdot \hat{e}_1) + V_2(\hat{e}_2 \cdot \hat{e}_1) \\
&= V_1\delta_{11} + V_2\delta_{21} \\
&= V_1. \tag{14.19}
\end{aligned}$$

This works out nicely because of the orthonormality of the basis vectors.

In the primed system, the same vector can be written in terms of its axis-parallel projected components and the primed basis vectors, as shown in Figure 14.3,

$$\overline{V} = V_1'\,\hat{g}_1' + V_2'\,\hat{g}_2'. \tag{14.20}$$

These primed vector components can also be determined from trigonometry, but because of the skewed geometry, this is no longer as simple as it was in the orthogonal system. Because the primed basis vectors are not orthogonal, an attempt to isolate a particular primed component by a vector manipulation similar to Equation 14.19 fails:

$$\begin{aligned}
\overline{V} \cdot \hat{g}_1' &= (V_1'\,\hat{g}_1' + V_2'\,\hat{g}_2') \cdot \hat{g}_1' \\
&= V_1'(\hat{g}_1' \cdot \hat{g}_1') + V_2'(\hat{g}_2' \cdot \hat{g}_1') \\
&= V_1' + V_2'(\hat{g}_2' \cdot \hat{g}_1') \\
&\neq V_1'. \tag{14.21}
\end{aligned}$$

It appears that vector manipulations in non-orthonormal coordinate systems are much more difficult than in orthonormal systems. Fortunately, there are some formal techniques which simplify this process. In the next section, we introduce the concepts of covariance, contravariance, and the metric tensor. Using these devices, the dot product between two vectors has the same form in an orthogonal system as in a non-orthogonal system.

14.2.2   Covariance, Contravariance, and the Metric

The basic complication introduced by a non-orthonormal coordinate system is evident in the dot product operation. In the two-dimensional orthonormal system described above, the inner product between two vectors is given by

$$\overline{A} \cdot \overline{B} = A_i \hat{e}_i \cdot B_j \hat{e}_j$$
$$= A_i B_j \delta_{ij}$$
$$= A_1 B_1 + A_2 B_2. \tag{14.22}$$

If this same inner product is performed in the non-orthogonal system of Figure 14.3, the result contains some extra terms:

$$\overline{A} \cdot \overline{B} = A_i' \, \hat{g}_i' \cdot B_j' \, \hat{g}_j'$$
$$= A_i' B_j' (\hat{g}_i' \cdot \hat{g}_j')$$
$$= A_1' B_1' + A_2' B_2' + (A_1' B_2' + A_2' B_1')(\hat{g}_1' \cdot \hat{g}_2'). \tag{14.23}$$

The inner product formed in the non-orthonormal system, as expressed in Equation 14.23, can be placed in the from of Equation 14.22 by rearranging as follows:

$$\overline{A} \cdot \overline{B} = A_1'[B_1' + B_2'(\hat{g}_1' \cdot \hat{g}_2')] + A_2'[B_1'(\hat{g}_1' \cdot \hat{g}_2') + B_2']. \tag{14.24}$$

Now define a new set of components for $\overline{B}$ as

$$\tilde{B}_1' = B_1' + B_2'(\hat{g}_1' \cdot \hat{g}_2')$$
$$\tilde{B}_2' = B_1'(\hat{g}_1' \cdot \hat{g}_2') + B_2'. \tag{14.25}$$

These quantities are called the *covariant* components of $\overline{B}$, while the original components are called *contravariant*. Clearly, the vector $\overline{B}$ cannot be expressed by combining these new covariant components with the basis vectors $\hat{g}_1'$ and $\hat{g}_2'$:

$$\overline{B} \neq \tilde{B}_1' \hat{g}_1' + \tilde{B}_2' \hat{g}_2'. \tag{14.26}$$

However, with these components, the inner product evaluated in the skewed system can be placed in the simple form

$$\overline{A} \cdot \overline{B} = A_i' \tilde{B}_i'$$
$$= A_1' \tilde{B}_1' + A_2' \tilde{B}_2'. \tag{14.27}$$

Notice that the inner product could also have been written as

$$\overline{A} \cdot \overline{B} = \tilde{A}_i' B_i', \tag{14.28}$$

with the covariant components of $\overline{A}$ defined as

$$\tilde{A}_1' = A_1' + A_2'(\hat{g}_1' \cdot \hat{g}_2')$$
$$\tilde{A}_2' = A_1'(\hat{g}_1' \cdot \hat{g}_2') + A_2'. \tag{14.29}$$

The inner product needs to be formed with mixed contravariant and covariant components, but it does not matter which vector is expressed with which type.

These arguments can be extended to non-orthogonal systems of arbitrary dimensions. The restriction that the basis vectors be normalized to unity can also be dropped. The covariant components can be generated from the contravariant components using the general expression

$$\tilde{A}_i' = A_j'(\hat{\mathbf{g}}_i' \cdot \hat{\mathbf{g}}_j'). \tag{14.30}$$

We have used the subscript summation convention here to imply a sum over the index j. For an n-dimensional coordinate system, there will be n terms in each sum. Notice that if the coordinate system is orthonormal, Equation 14.30 reduces to $\tilde{A}_i' = A_i'$, and the covariant and contravariant components are equal. In this case, both Equations 14.27 and 14.28 will revert back to Equation 14.22. This is important, because it implies this new notation is general enough to handle all of our previous Cartesian and curvilinear coordinate systems, as well as the new non-orthonormal ones.

There is another way of expressing the inner product between two vectors in a non-orthonormal system that makes use of a quantity called *the metric*. As will be shown later, the metric turns out to be a second-rank tensor. The elements of the metric, in an unprimed system, are defined as

$$M_{ij} \equiv \hat{\mathbf{g}}_i \cdot \hat{\mathbf{g}}_j. \tag{14.31}$$

Notice this definition implies that the metric is symmetric:

$$M_{ij} = M_{ji}. \tag{14.32}$$

Using the metric, Equation 14.30 can be written as

$$\tilde{A}_i = A_j M_{ij}. \tag{14.33}$$

The metric converts the contravariant components to covariant components.

Now the inner product between $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ can be written as

$$\overline{\mathbf{A}} \cdot \overline{\mathbf{B}} = A_i B_j M_{ij}. \tag{14.34}$$

A sum over both i and j is indicated on the RHS of this equation. If the sum over i is performed first, Equation 14.34 becomes

$$\overline{\mathbf{A}} \cdot \overline{\mathbf{B}} = \tilde{A}_j B_j. \tag{14.35}$$

When the sum over j is performed first, Equation 14.34 instead becomes

$$\overline{\mathbf{A}} \cdot \overline{\mathbf{B}} = A_i \tilde{B}_i. \tag{14.36}$$

When Equation 14.34 is used for the inner product, the vector components are not mixed. Contravariant components are used for both. If the system is orthonormal, $M_{ij} = \delta_{ij}$, and the standard inner product for orthonormal systems results. Notice

that the metric is determined solely by the basis vectors of the coordinate system. This turns out to be an important fact, and will allow us to identify the metric as a second-rank tensor.

In summary, there are two ways to perform the inner product between two vectors in a non-orthonormal system. One way is to use the covariant and contravariant components, as was done in Equations 14.27 and 14.28. A completely equivalent method is to use the metric and the regular contravariant vector components, as demonstrated in Equation 14.34. These arguments can be naturally extended to inner products between tensor quantities, but this generalization will be postponed until the transformation equations for non-orthonormal systems are worked out.

14.2.3 Contravariant Vector Component Transformations

Imagine two different skewed coordinate systems, as shown in Figure 14.4. We want to find out how the contravariant components of a vector expressed in the first system can be transformed into the second system. The first system has the unprimed coordinates x_i and the basis vectors $\hat{g}_i$, while the second system uses the primed coordinates x_i' and basis vectors $\hat{g}_i'$. Remember, we are still limiting ourselves to coordinate systems with constant basis vectors. Let the general equations which relate the two sets of coordinates to one another be

$$x_i' = x_i'(x_1, x_2, x_3)$$
$$x_i = x_i(x_1', x_2', x_3').$$ (14.37)

There will be one pair of equations for each dimension of the systems.

In our previous work dealing with transformations between orthonormal coordinate systems, we were able to relate vector components of one system to another via
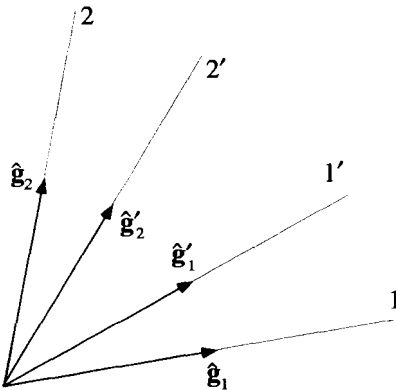
Figure 14.4 Two Skewed Coordinate Systems

[a], the transformation matrix:

$$V_i' = a_{ij} V_j. \tag{14.38}$$

The restriction to orthonormal systems allowed us to invert this expression easily, because it turned out that $a_{ij}^{-1} = a_{ji}$. We can still write a relationship similar to Equation 14.38 for transformations between non-orthonormal systems, but we need to be more careful, because the inverse of the transformation matrix is no longer simply its transpose. To keep tabs on which transformations have this simple inversion and which do not, we will reserve the matrix [a] for transformations between orthonormal systems. The matrix [t] will represent transformations from unprimed coordinates to primed coordinates, where the systems may be non-orthonormal:

$$V_i' = t_{ij} V_j. \tag{14.39}$$

The reverse operation, a transformation from primed coordinates to unprimed coordinates, will use the matrix [g],

$$V_i = g_{ij} V_j', \tag{14.40}$$

where $g_{ij} = t_{ij}^{-1} \neq t_{ji}$. By their very definition, it follows that $t_{ij} g_{jk} = \delta_{ik}$. We will discuss the relationship between the [t] and [g] matrices in more detail later. In both these expressions, the vector components are the regular, contravariant components of $\overline{V}$, not the covariant components we introduced earlier.

All the vectors at a given point transform using the same [t] matrix. To determine the t_{ij} elements, it is easiest to consider the displacement vector $d\overline{r}$, which in both coordinate systems is given by

$$d\overline{r} = dx_i \, \hat{\mathbf{g}}_i = dx_i' \, \hat{\mathbf{g}}_i'. \tag{14.41}$$

Applying this equality to Equation 14.39 gives

$$dx_i' = t_{ij} \, dx_j. \tag{14.42}$$

Referring to Equations 14.37 gives the relationship

$$dx_i' = \frac{\partial x_i'}{\partial x_j} \, dx_j, \tag{14.43}$$

and the transformation matrix elements can be written as

$$t_{ij} = \frac{\partial x_i'}{\partial x_j}. \tag{14.44}$$

So far, these results look very similar to those of the Cartesian transformations of Chapter 4. In fact, the equation for the components of [t] given in Equation 14.44 is the same result obtained for the [a] matrix between Cartesian systems. The complications

arise when we try to invert these equations. As we mentioned before, the inversion of $[t]$ is no longer simply its transpose. A general way to obtain $[t]^{-1}$, what we are calling $[g]$, is to use the expression

$$g_{ij} \equiv t_{ij}^{-1} = \frac{c_{ji}}{|t_{ij}|_{det}}, \tag{14.45}$$

where c_{ji} is the ji-cofactor of the t_{ij} matrix. From matrix algebra, this cofactor is defined as $(-1)^{i+j}$ times the determinant of the t_{ij} matrix, with the jth row and the ith column removed. The $[g]$ matrix can also be obtained from the equations relating the coordinates in exactly the same manner that led to Equation 14.44:

$$g_{ij} \equiv t_{ij}^{-1} = \frac{\partial x_i}{\partial x_j'}. \tag{14.46}$$

The $[t]$ and $[g]$ matrices can also be used to relate the basis vectors to one another. Using the contravariant components, any vector $\overline{V}$ can be expressed in the primed or unprimed systems as

$$\overline{V} = V_j\,\hat{\mathbf{g}}_j = V_i'\,\hat{\mathbf{g}}_i'. \tag{14.47}$$

Substituting Equation 14.39 into Equation 14.47 gives

$$\overline{V} = V_j\,\hat{\mathbf{g}}_j = V_j\,t_{ij}\,\hat{\mathbf{g}}_i'. \tag{14.48}$$

Since this expression must be true for any $\overline{V}$, we must have

$$\hat{\mathbf{g}}_j = t_{ij}\,\hat{\mathbf{g}}_i'. \tag{14.49}$$

Following the same steps above, using $[g]$ instead of $[t]$ gives

$$\hat{\mathbf{g}}_j' = g_{ij}\,\hat{\mathbf{g}}_i. \tag{14.50}$$

Notice the contravariant vector components are transformed by contractions over the second subscript of either t_{ij} or g_{ij}, while the basis vectors are transformed by contracting over the first subscript.

To summarize the results of this section, the transformation between the two non-orthonormal coordinate systems is governed by the relations

$t_{ij} = \partial x_i'/\partial x_j$	$g_{ij} = \partial x_i/\partial x_j'$
$V_i' = t_{ij}V_j$	$V_i = g_{ij}V_j'$
$\hat{\mathbf{g}}_j = t_{ij}\,\hat{\mathbf{g}}_i'$	$\hat{\mathbf{g}}_j' = g_{ij}\,\hat{\mathbf{g}}_i.$

14.2.4 Subscript/Superscript Notation

Before proceeding with a discussion of how covariant vector components transform, it turns out to be convenient to introduce some new notation. The tilde ($\tilde{V}_i$) notation we have been using for the covariant vector components is clumsy. It is not obvious that the following conventions are much better, but they do provide a valuable mechanism for keeping track of which type of component (contravariant or covariant) should be used in an expression. The standard, axis-parallel projected vector components, which we have called the contravariant components, will now be labeled with a superscript, while the new covariant components will use a subscript instead of a tilde. For example, the contravariant components of the vector $\overline{V}$ are V^i, while the covariant components are V_i.

One advantage of this new notation is evident by looking at the form of the inner product. With the superscript/subscript convention, we can write the dot product of $\overline{A}$ and $\overline{B}$ as

$$\overline{A} \cdot \overline{B} = A^i B_i = A_i B^i. \tag{14.51}$$

Notice the index being summed over appears once as a subscript and once as a superscript. This, of course, is the same as saying that the sum is done over mixed contravariant and covariant quantities. This process of mixed superscripts and subscripts will persist for almost all contractions over a repeated index. It even works in forming a vector from its components with the proper interpretation of the basis vectors. We know that the vector can be formed with the contravariant vector components and the basis vectors:

$$\overline{V} = V^i \hat{g}_i. \tag{14.52}$$

To be consistent with the subscript/superscript convention, the basis vectors must be labeled with subscripts and be considered covariant. We will see, in the next section, that this conclusion is consistent with the way these basis vectors transform.

This convention also prevents us from accidently forming a vector by combining its covariant vector components with the $\hat{g}_i$ basis vectors:

$$\overline{V} \neq V_i \hat{g}_i. \tag{14.53}$$

The notation warns us that this is not correct because both indices appear as subscripts.

In the previous section we generated several relations which described how the contravariant components of a vector $\overline{V}$ transform between two skewed systems. How should the presentation of these results be modified to be consistent with the new superscript/subscript convention? In the previous section we had written

$$V_i' = t_{ij} V_j. \tag{14.54}$$

Now these vector components need to be superscripted. To be consistent with this new notation, one of the indices of the transformation matrix needs to be a subscript

and one a superscript,

$$V'^i = t^i{}_j V^j, \tag{14.55}$$

where

$$t^i{}_j = \frac{\partial x'^i}{\partial x^j}. \tag{14.56}$$

In a similar manner, the inversion of Equation 14.55 becomes

$$V^i = g^i{}_j V'^j, \tag{14.57}$$

where

$$g^i{}_j = \frac{\partial x^i}{\partial x'^j}. \tag{14.58}$$

In Equations 14.56 and 14.58, notice how the superscripted component in the denominator of the partial derivative results in a subscripted index on the LHS of this expression. This is a general property of partial derivatives with respect to contravariant and covariant quantities. A partial derivative with respect to a contravariant quantity produces a covariant result, while a partial derivative with respect to a covariant quantity gives a contravariant result. We will prove this fact later in the chapter.

These transformation matrices have what is called *mixed* contravariant/covariant properties. They are contravariant with respect to one index, but covariant with respect to another. We have not really defined what these words mean for quantities with more than one index yet, and we will defer this to a later section of this chapter, where we talk about tensor transformations.

With the earlier notation, the reciprocal nature of [t] and [g] was indicated by the equation $t_{ij} g_{jk} = \delta_{ik}$. But now, to be consistent with the subscript/superscript convention, we should write

$$g^i{}_j t^j{}_k = \delta^i{}_k. \tag{14.59}$$

The Kronecker-δ, written in this way, also has mixed contravariant and covariant form.

Equations 14.49 and 14.50, which indicate how the basis vectors transform, are written using superscript/subscript notation as

$$\hat{\mathbf{g}}_j = t^i{}_j \, \hat{\mathbf{g}}'_i$$

$$\hat{\mathbf{g}}'_j = g^i{}_j \, \hat{\mathbf{g}}_i. \tag{14.60}$$

Notice how the horizontal positioning of the indices of the transformation matrix is important. In Equation 14.55 and 14.57, the sum was over the second index of the matrix, while these sums are over the first index. This prevents us from writing the [t] matrix elements as t^i_j, since this would no longer indicate which index comes first.

We should also rewrite the relations involving the metric using the new notation. Our previous definition of the metric was in terms of the covariant basis vectors. Consequently, Equation 14.31 remains unchanged:

$$M_{ij} = \hat{\mathbf{g}}_i \cdot \hat{\mathbf{g}}_j, \tag{14.61}$$

and both indices remain as subscripts. Formed this way, the metric elements are purely covariant because both indices are subscripts. The metric converts the contravariant components of a vector to its covariant components, within the same coordinate system. This operation can be written, using the superscript/subscript notation, as

$$V_i = M_{ij} V^j. \tag{14.62}$$

Notice how the summation convention continues to work. This same operation in a primed system uses a primed metric,

$$M'_{ij} = \hat{\mathbf{g}}'_i \cdot \hat{\mathbf{g}}'_j, \tag{14.63}$$

and is written as

$$V'_i = M'_{ij} V'^j. \tag{14.64}$$

In summary, the equations governing the transformations of the contravariant components of a vector $\overline{\mathbf{V}}$ can be written using the new superscript/subscript notation as:

$$\begin{array}{ll}
t^i{}_j = \partial x'^i / \partial x^j & g^i{}_j = \partial x^i / \partial x'^j \\[2mm]
V'^i = t^i{}_j V^j & V^i = g^i{}_j V'^j \\[2mm]
\hat{\mathbf{g}}_j = t^i{}_j \hat{\mathbf{g}}'_i & \hat{\mathbf{g}}'_j = g^i{}_j \hat{\mathbf{g}}_i.
\end{array}$$

The covariant components of $\overline{\mathbf{V}}$ can be obtained from the contravariant components using the metric:

$$\begin{array}{ll}
M_{ij} = \hat{\mathbf{g}}_i \cdot \hat{\mathbf{g}}_j & V_i = M_{ij} V^j \\[2mm]
M'_{ij} = \hat{\mathbf{g}}'_i \cdot \hat{\mathbf{g}}'_j & V'_i = M'_{ij} V'^j.
\end{array}$$

There are clearly some holes in this picture. First, there is the question of how the covariant components of a vector transform. Second, we said the basis vectors $\hat{\mathbf{g}}_i$ were covariant in nature. We need to prove this. Finally, can we define contravariant basis vectors? These issues are all intimately related to one another, and are addressed in the next section.

14.2.5   Covariant Vector Component Transformations

Return to the pair of skewed coordinate systems described in Figure 14.4. The covariant vector components of the vector $\overline{V}$ will transform according to some linear relation

$$V_i' = [?]V_j. \tag{14.65}$$

To determine [?] in this expression, consider two equivalent forms of the inner product of two vectors, one in the primed system and one in the unprimed system:

$$\overline{A} \cdot \overline{B} = A^i B_i = A'^j B_j'. \tag{14.66}$$

The contravariant components of $\overline{A}$ transform according to the rules determined earlier:

$$A'^j = t^j_i A^i. \tag{14.67}$$

Substituting this expression into the far RHS of Equation 14.66 gives

$$A^i B_i = A^i t^j_i B_j'. \tag{14.68}$$

Since this equation must be true for any $\overline{A}$, we must have

$$B_i = t^j_i B_j'. \tag{14.69}$$

This expression is easily inverted to give

$$B_i' = g^j_i B_j. \tag{14.70}$$

Notice the similarity between Equations 14.60, 14.69, and 14.70. This supports our conclusion that the axis-parallel basis vectors are covariant.

We were able to combine the contravariant components of a vector with the covariant basis vectors to form the vector itself:

$$\overline{V} = V^i \hat{\mathbf{g}}_i. \tag{14.71}$$

It would be nice to come up with a new set of contravariant basis vectors, $\hat{\mathbf{g}}^i$, which could be combined with the covariant vector component to form the same vector. That is,

$$\overline{V} = V_i \hat{\mathbf{g}}^i. \tag{14.72}$$

In fact, we can use this expression to define these contravariant basis vectors, and look at the consequences.

The basic properties of the contravariant basis vectors can be deduced by again considering the inner product between two vectors, $\overline{A}$ and $\overline{B}$. If $\overline{A}$ is expressed using covariant basis vectors and contravariant components, while $\overline{B}$ is written with con-

travariant basis vectors and covariant vector components, the inner product becomes

$$\overline{\mathbf{A}} \cdot \overline{\mathbf{B}} = A^i \, \hat{\mathbf{g}}_i \cdot B_j \, \hat{\mathbf{g}}^j$$
$$= A^i \, B_j \, \hat{\mathbf{g}}_i \cdot \hat{\mathbf{g}}^j. \qquad (14.73)$$

According to Equation 14.51, this expression must be equal to $A^i \, B_i$ and so

$$\hat{\mathbf{g}}_i \cdot \hat{\mathbf{g}}^j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}, \qquad (14.74)$$

or in terms of the Kronecker symbol,

$$\hat{\mathbf{g}}_i \cdot \hat{\mathbf{g}}^j = \delta_i{}^j. \qquad (14.75)$$

This last condition allows both the magnitude and direction of the contravariant basis vectors to be determined, if the covariant basis vectors are already known. Working in two dimensions, $\hat{\mathbf{g}}^1 \cdot \hat{\mathbf{g}}_2 = 0$ and $\hat{\mathbf{g}}^1 \cdot \hat{\mathbf{g}}_1 = 1$. In words, $\hat{\mathbf{g}}^1$ must be perpendicular to $\hat{\mathbf{g}}_2$, while its projection along the 1-axis, parallel to $\hat{\mathbf{g}}_1$, must be one. This uniquely determines $\hat{\mathbf{g}}^1$ and, by similar arguments, $\hat{\mathbf{g}}_2$. The conditions of Equation 14.75 can be pictured graphically as shown in Figure 14.5. The constructions in this figure have been made assuming that $|\hat{\mathbf{g}}_i| = 1$.

The covariant and contravariant vector components can also be interpreted graphically, as shown in Figure 14.6. Again, for this figure, it has been assumed that $|\hat{\mathbf{g}}_i| = 1$. The contravariant vector components are simply the magnitudes of the axis-parallel projections of the vector onto the skewed axes defined by the covariant basis vectors. The covariant components are the magnitudes of the projections of the vector onto the same coordinate axes, but following lines parallel to the new contravariant basis vectors. This makes the projection lines for the covariant vector components
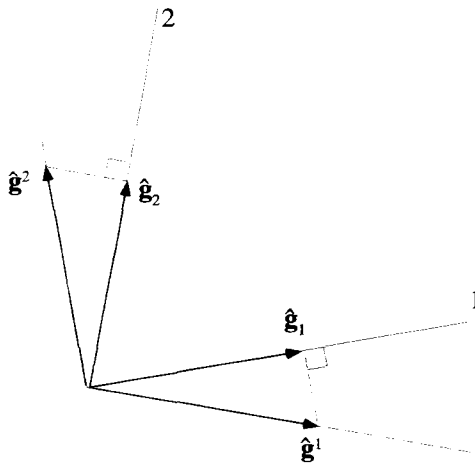
Figure 14.5 Determination of the Contravariant Basis Vectors
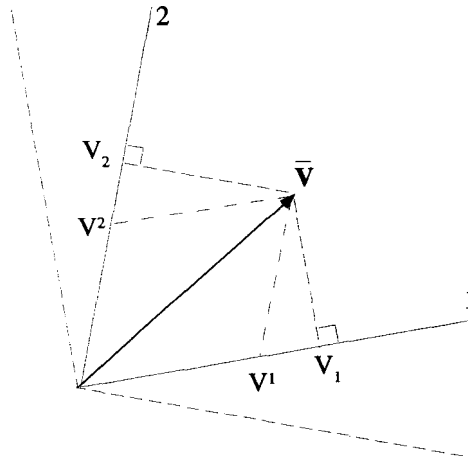
Figure 14.6   Covariant and Contravariant Component Projections of a Vector

perpendicular to the axes, as shown in this figure. The geometry insures that

$$\overline{\mathbf{V}} = V^i\,\hat{\mathbf{g}}_i = V_i\,\hat{\mathbf{g}}^i. \tag{14.76}$$

If the covariant basis vectors are not unit vectors, the constructions of Figures 14.5 and 14.6 must be adjusted appropriately, following the requirements of Equations 14.75 and 14.76.

Transformations for the contravariant basis vectors fall directly from Equation 14.76 by the techniques we have applied several times:

$$\hat{\mathbf{g}}^{\prime i} = t^i{}_j\,\hat{\mathbf{g}}^j \tag{14.77}$$

$$\hat{\mathbf{g}}^i = g^i{}_j\,\hat{\mathbf{g}}^{\prime j}. \tag{14.78}$$

This confirms our classification of these new basis vectors as contravariant, because they transform exactly like the contravariant components of a vector.

The complete set of transformation rules for both contravariant and covariant vector components and basis vectors can be summarized by the symmetric set of relations:

$$
\begin{array}{ll}
V^{\prime i} = t^i{}_j\,V^j & \hat{\mathbf{g}}^{\prime i} = t^i{}_j\,\hat{\mathbf{g}}^j \\
V^i = g^i{}_j\,V^{\prime j} & \hat{\mathbf{g}}^i = g^i{}_j\,\hat{\mathbf{g}}^{\prime j} \\[2mm]
V_i^{\prime} = g^j{}_i\,V_j & \hat{\mathbf{g}}_i^{\prime} = g^j{}_i\,\hat{\mathbf{g}}_j \\
V_i = t^j{}_i\,V_j^{\prime} & \hat{\mathbf{g}}_i = t^j{}_i\,\hat{\mathbf{g}}_j^{\prime}
\end{array}
$$

with

$$t^i{}_j = \partial x^{\prime i}/\partial x^j \qquad g^i{}_j = \partial x^i/\partial x^{\prime j}.$$

Notice that contravariant quantities are always transformed by a summation over the second index of either $t^i_{\;j}$ or $g^i_{\;j}$, while covariant quantities transform by summing over the first index. For contravariant quantities, $t^i_{\;j}$ is used to go from the unprimed to the primed system, while $g^i_{\;j}$ is used to go from the primed to the unprimed system. For covariant quantities, the roles of $t^i_{\;j}$ and $g^i_{\;j}$ are reversed.

The new contravariant basis vectors allow us to construct another version of the metric, this time with superscripts:

$$M^{ij} = \hat{\mathbf{g}}^i \cdot \hat{\mathbf{g}}^j. \tag{14.79}$$

Application of this form of the metric converts covariant quantities into contravariant quantities. For example,

$$V^i = M^{ij}\, V_j. \tag{14.80}$$

We will see in the next section that the two different metrics, M_{ij} and M^{ij}, are simply different representations of the same object, the metric tensor.

14.2.6 Covariance and Contravariance in Tensors

The covariant and contravariant properties discussed above are easily extended to tensors. Just as a vector can be expressed with contravariant or covariant components,

$$\overline{\mathbf{V}} = V^i\, \hat{\mathbf{g}}_i = V_i\, \hat{\mathbf{g}}^i, \tag{14.81}$$

a tensor can be expressed using purely contravariant or covariant components:

$$\overline{\overline{\mathbf{T}}} = T^{ijk}\, \hat{\mathbf{g}}_i\, \hat{\mathbf{g}}_j\, \hat{\mathbf{g}}_k = T_{ijk}\, \hat{\mathbf{g}}^i\, \hat{\mathbf{g}}^j\, \hat{\mathbf{g}}^k. \tag{14.82}$$

However, higher-rank tensors are more flexible than vectors because they can also be expressed in a mixed form, with both contravariant and covariant indices. For example,

$$\overline{\overline{\mathbf{T}}} = T^{i\;k}_{\;j}\, \hat{\mathbf{g}}_i\, \hat{\mathbf{g}}^j\, \hat{\mathbf{g}}_k \tag{14.83}$$

is another equivalent representation of $\overline{\overline{\mathbf{T}}}$.

All the tensor expressions in Equations 14.82 and 14.83 are equivalent, although the specific values of the components will be different in each case. Just as the covariant and contravariant components of a vector are related via the metric, the different representations of $\overline{\overline{\mathbf{T}}}$ can be obtained from one another using the same metric. For example, if the two expressions for $\overline{\overline{\mathbf{T}}}$ in Equation 14.82 are equal, we can write

$$T^{ijk} = M^{il}\, M^{jm}\, M^{kn}\, T_{lmn}. \tag{14.84}$$

The expression in Equation 14.83 gives the same tensor when

$$T^{i\,k}_{\ j} = M_{jm}\, T^{imk}. \tag{14.85}$$

To convert a set of tensor components from purely covariant to purely contravariant form, one metric operation is needed for each index.

Coordinate system transformations of tensors follow the same pattern we established for vector transformations. One transformation matrix of the appropriate kind is used for each index. For example, we might write

$$T'^{\ j}_{i\,k} = g^{l}_{\ i}\, t'^{j}_{\ m}\, g^{n}_{\ k}\, T_{l}^{\ m}_{\ n} \tag{14.86}$$

$$T_{i\,k}^{\ j} = t^{l}_{\ i}\, g^{j}_{\ m}\, t^{n}_{\ k}\, T'^{\ m}_{l\ n}. \tag{14.87}$$

Example 14.2   Reference has been made to the fact that the metric is a tensor, but we have not really proven this fact. The proof is straightforward. Consider the metric elements, expressed in pure covariant form as,

$$M_{ij} = \hat{\mathbf{g}}_i \cdot \hat{\mathbf{g}}_j. \tag{14.88}$$

The inner product between two vectors, expressed in two different coordinate systems, can be written

$$\overline{\mathbf{A}} \cdot \overline{\mathbf{B}} = A^i\, B^j\, M_{ij} = A'^{m}\, B'^{n}\, M'_{mn}, \tag{14.89}$$

where $M'_{mn} = \hat{\mathbf{g}}'_m \cdot \hat{\mathbf{g}}'_n$. The transformation equations can be used to express the primed vector components in terms of the unprimed components. This gives

$$A^i\, B^j\, M_{ij} = A^i\, B^j\, t^{m}_{\ i}\, t^{n}_{\ j}\, M'_{mn}. \tag{14.90}$$

Since this expression must work for all possible $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$, we must have

$$M_{ij} = t^{m}_{\ i}\, t^{n}_{\ j}\, M'_{mn}, \tag{14.91}$$

which easily inverts to give

$$M'_{ij} = g^{m}_{\ i}\, g^{n}_{\ j}\, M_{mn}. \tag{14.92}$$

But this is exactly how the elements of a second-rank tensor must transform, and so by definition the metric is a tensor. This means that we can write

$$\overline{\overline{\mathbf{M}}} = M_{ij}\, \hat{\mathbf{g}}^i\, \hat{\mathbf{g}}^j. \tag{14.93}$$

Since the metric is a tensor, we can modify its covariant or contravariant nature as we would for any tensor. Although it might seem a little odd to use the metric to modify itself, we can change a purely covariant metric to a purely contravariant form by applying the metric twice:

$$M^{ij} = M^{im} M^{jn} M_{mn}. \tag{14.94}$$

We can also put the metric in mixed form by writing

$$M^i{}_j = M^{im} M_{mj}. \tag{14.95}$$

Using the transformation equations, you can easily show that

$$M^i{}_j = \hat{\mathbf{g}}^i \cdot \hat{\mathbf{g}}_j = \delta^i{}_j. \tag{14.96}$$

This implies that the metric tensor is really just a generalization of the Kronecker-δ tensor.

14.2.7   The Contravariance and Covariance of Partial Derivatives

When partial derivatives are taken with respect to a contravariant coordinate, the result is a covariant quantity. To see this, let the contravariant coordinates in an arbitrary pair of coordinate systems be x^i and x'^i. The rules of calculus require that

$$\frac{\partial}{\partial x'^i} = \frac{\partial x^j}{\partial x'^i} \frac{\partial}{\partial x^j}, \tag{14.97}$$

where there is an implied summation over the index j. But notice, the $\partial x^j / \partial x'^i$ term is exactly the definition of $g^j{}_i$. This lets us write

$$\frac{\partial}{\partial x'^i} = g^j{}_i \frac{\partial}{\partial x^j}. \tag{14.98}$$

Comparing this expression with Equation 14.70 shows that the partial derivative operation transforms just like a covariant quantity. The same type of argument holds for partial derivatives with respect to covariant coordinates. In that case we find

$$\frac{\partial}{\partial x'_i} = \frac{\partial x_j}{\partial x'_i} \frac{\partial}{\partial x_j} \tag{14.99}$$

$$= t^j{}_i \frac{\partial}{\partial x_j}, \tag{14.100}$$

which says that this partial derivative acts like a contravariant quantity. To be consistent with our superscript/subscript conventions, we impose the rule that a superscript in the "denominator" of the derivative operation acts like a subscript, while a subscript in the denominator acts like a superscript. This idea was discussed briefly in connection with the transformation matrices in Equations 14.56 and 14.58.

Example 14.3   A static electric field is often calculated by taking the gradient of a scalar potential:

$$\overline{\mathbf{E}} = -\overline{\nabla} \phi. \tag{14.101}$$

In an earlier chapter, we defined the gradient operator by the relation

$$d\phi = \overline{\boldsymbol{\nabla}}\phi \cdot d\overline{\mathbf{r}}. \tag{14.102}$$

Since the displacement vector can be written

$$d\overline{\mathbf{r}} = dx^i \, \hat{\mathbf{g}}_i, \tag{14.103}$$

where dx^i is a contravariant quantity, it is clear the gradient of ϕ can be written as

$$\overline{\boldsymbol{\nabla}}\phi = \frac{\partial\phi}{\partial x^i} \, \hat{\mathbf{g}}^i. \tag{14.104}$$

To check the validity of this expression, insert Equations 14.104 and 14.103 into the RHS of Equation 14.102 to obtain

$$\overline{\boldsymbol{\nabla}}\phi \cdot d\overline{\mathbf{r}} = \frac{\partial\phi}{\partial x^i} \, \hat{\mathbf{g}}^i \cdot dx^j \, \hat{\mathbf{g}}_j$$

$$= \frac{\partial\phi}{\partial x^i} \, dx^j \, \delta^i{}_j \tag{14.105}$$

$$= \frac{\partial\phi}{\partial x^i} \, dx^i$$

$$= d\phi. \tag{14.106}$$

When we write the components of the electric field as

$$E_i = -\frac{\partial\phi}{\partial x^i}, \tag{14.107}$$

they are covariant and must transform according to the relation

$$E_i' = g^j{}_i E_j. \tag{14.108}$$

Example 14.4 A static magnetic field $\overline{\mathbf{B}}$ can be calculated using Ampere's law:

$$\oint_C d\overline{\mathbf{r}} \cdot \overline{\mathbf{B}} = \mu_o I. \tag{14.109}$$

In this expression, I is the total current flowing through the closed path C. Taking $d\overline{\mathbf{r}}$ to be a differential vector quantity with contravariant components, as given in Equation 14.103, the components of the magnetic field used in this integration should be written in covariant form, so that

$$\oint_C dx^i \, B_i = \mu_o I. \tag{14.110}$$

14.2.8 Special Relativity

Special relativity deals with the observation of events as seen from two different coordinate systems, one moving at constant velocity with respect to the other. For simplicity, we will limit most of this discussion to systems with one spatial dimension, as shown in Figure 14.7. The primed system is moving, relative to the unprimed system, with a velocity v_o in the x-direction. Einstein's favorite example of this situation was a railway car moving at a constant velocity on straight, flat tracks. A coordinate system fastened to the tracks, which he called the embankment, is the unprimed system and a system attached to the moving car is the primed system. These two coordinate systems form different reference frames for the observation of events.

Special relativity treats time as another dimension, on an equal footing with the spatial coordinates. To make this work out, the time coordinate needs to have the same units as the spatial coordinate. This is accomplished by multiplying the time parameters by the speed of light c_o. An event occurs at a specific set of values of the space-time coordinates. For these coordinates in the unprimed system, instead of x^1 and x^2, we will use x and $c_o t$:

$$x^1 \rightarrow x$$
$$x^2 \rightarrow c_o t.$$

Even though we have not labeled these coordinates with superscripts, you should realize that they are actually contravariant quantities. In the primed system, this same event occurs at the primed coordinates:

$$x'^1 \rightarrow x'$$
$$x'^2 \rightarrow c_o t'.$$

The Lorentz Transformation   In the previous section, we showed two different sets of coordinates could describe the same event. Now we ask, what is the transformation that converts one set of these coordinates to another? To answer this, we need to briefly discuss the fundamental assumption of special relativity.

The basic postulate of special relativity is that the speed of light must be constant, regardless of the reference frame from which it is being measured. This is a strange idea, apparently fraught with paradox. It implies, for example, that the light coming out of the headlights of a moving car, as observed by someone standing on the
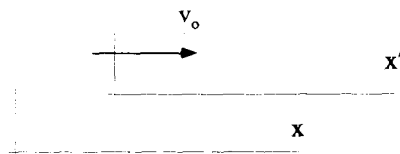
Figure 14.7   The Inertial Systems for Special Relativity

ground outside the car, travels at the same speed, independent of the speed of the car. It also has other strange consequences, such as the dilation of time and contraction of lengths for fast-moving objects. Nevertheless, it has been confirmed by many different experiments. Our goal here is not to delve into these ideas and their associated paradoxes, but rather to show how the transformation between inertial frames can be looked at using a skewed coordinate system. For a complete introduction to the ideas of relativity, the reader should investigate the excellent book, *Spacetime Physics*, by Edwin Taylor and John Wheeler.

What does the invariance of the speed of light say about the transformation matrix that converts the unprimed coordinates to the primed coordinates? In terms of the coordinates defined above, if an object travels at the speed of light, following the line

$$x = c_o t, \tag{14.111}$$

then the transformation matrix $[t]$ must convert this into

$$x' = c_o t', \tag{14.112}$$

so the object is still moving at the same speed in the primed coordinate system. The linear transformation that accomplishes this was worked out by Hendrick Lorentz, and bears his name. With one spatial dimension, the Lorentz transformation is given in matrix form as

$$t^i{}_j \rightarrow \begin{bmatrix} \gamma_o & -\gamma_o \beta_o \\ -\gamma_o \beta_o & \gamma_o \end{bmatrix}, \tag{14.113}$$

where

$$\beta_o = \frac{v_o}{c_o} \tag{14.114}$$

and

$$\gamma_o^2 = \frac{1}{1 - \beta_o^2}. \tag{14.115}$$

The inversion of the matrix in Equation 14.113 is straightforward and gives the elements of the $[g]$ matrix as

$$g^i{}_j \rightarrow \begin{bmatrix} \gamma_o & \gamma_o \beta_o \\ \gamma_o \beta_o & \gamma_o \end{bmatrix}. \tag{14.116}$$

It should not be surprising that this matrix is simply the $[t]$ matrix with the sign of v_o reversed.

The form of these transformations implies an important fact. Given the coordinates of an event observed in the unprimed frame $(x, c_o t)$, the quantity $x^2 - (c_o t)^2$ is invariant to transformation. That is, if we form this same quantity using the event's primed

components, we find that

$$x^2 - (c_o t)^2 = x'^2 - (c_o t')^2. \tag{14.117}$$

This statement is true for any x and $c_o t$, as long as the primed and unprimed coordinates are related by the Lorentz transformation. This is easily shown by substituting

$$x' = \gamma_o x - \gamma_o \beta_o c_o t \tag{14.118}$$

and

$$c_o t' = -\gamma_o \beta_o x + \gamma_o c_o t \tag{14.119}$$

into the RHS of Equation 14.117.

Skewed Coordinate Representation There is a clever way, using a skewed coordinate system, to visualize the relationship between the primed and unprimed coordinates that is imposed by the Lorentz transformation. Imagine the unprimed coordinate axes are perpendicular, and introduce the orthonormal, covariant basis vectors $\hat{g}_1$ and $\hat{g}_2$, as shown in Figure 14.8. The covariant basis vectors of the primed system can be determined using Equation 14.60 and the elements of the $g^i{}_j$ matrix:

$$\begin{bmatrix} \hat{g}'_1 & \hat{g}'_2 \end{bmatrix} = \begin{bmatrix} \hat{g}_1 & \hat{g}_2 \end{bmatrix} \begin{bmatrix} \gamma_o & \gamma_o \beta_o \\ \gamma_o \beta_o & \gamma_o \end{bmatrix}. \tag{14.120}$$

Notice we have arranged the basis vectors horizontally to be consistent with the fact that the summation is done over the first index of $g^i{}_j$. According to these equations, the primed basis vectors form a skewed set of coordinate axes, with an angle of $\theta = \tan^{-1} \beta_o$ between the primed and corresponding unprimed axis. This is shown in Figure 14.8. As the relative velocity between the reference frames increases, θ increases and the prime system becomes more drastically skewed.

This representation has several things going for it. An event, with coordinates $(x, c_o t)$ in the unprimed system and $(x', c_o t')$ in the primed system, can be represented as a single point for both systems, as shown in Figure 14.9. Except for a small complication due to how the axes are scaled, the contravariant components in each coordinate system can be determined by axis-parallel projections of the point onto the appropriate axes. Notice, the line $x = c_o t$, called a light line, plots as the straight line at $45°$. The same line could also be written as $x' = c_o t'$, graphically demonstrating that the speed of light is invariant.

If an object is observed in the unprimed frame, moving at constant velocity $x = vt$, the motion plots as a straight line in the two-dimensional coordinate systems. If $v < c_o$, this line must lie above the light line, as shown in Figure 14.10. The same line describes the motion as observed in the primed system, and so the motion in this frame is still at constant velocity, $x' = v't'$. Notice, however, that $v' \neq v$. We can
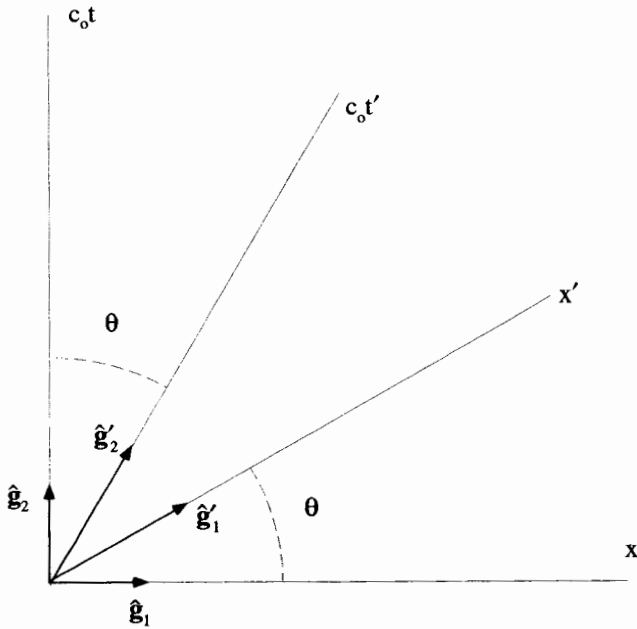
Figure 14.8 The Two-Dimensional Lorentz Coordinate Systems
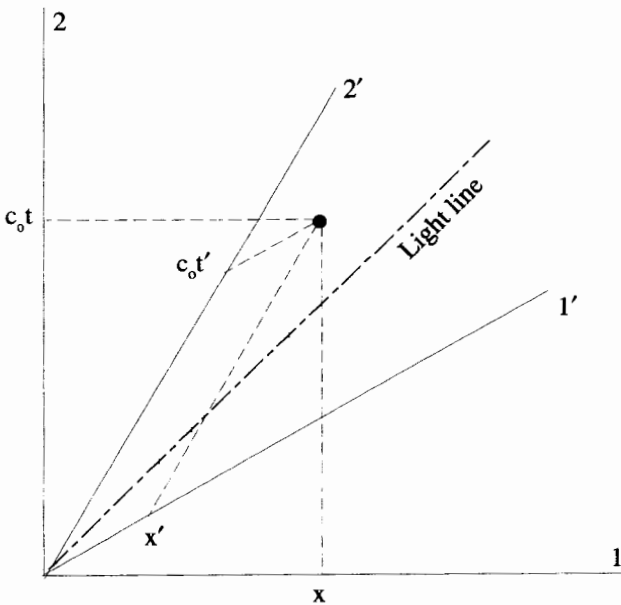
Figure 14.9 Lorentz Transformations of Events
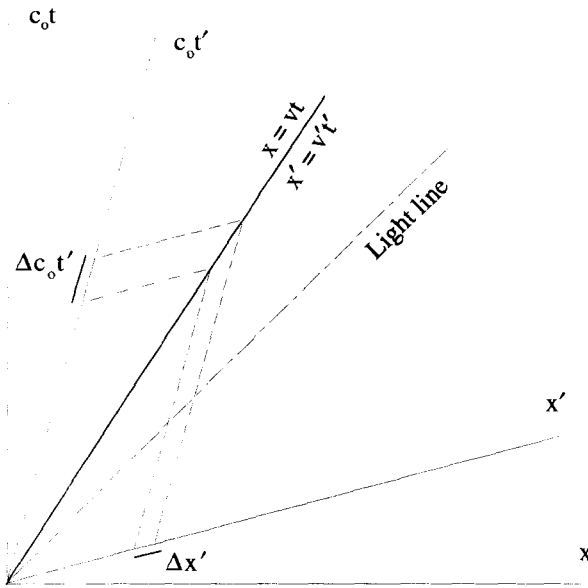
Figure 14.10   Motion at a Velocity Less Than the Speed of Light

determine the primed velocity from:

$$v' = \frac{\Delta x'}{\Delta(c_o t')}.$$ (14.121)

This velocity can be calculated directly from the transformation equations as

$$v' = c_o \left(\frac{v - \beta_o c_o}{c_o - \beta_o v} \right).$$ (14.122)

As can be seen from this expression and from the figure, as $v \rightarrow c_o$, $v' \rightarrow c_o$.

Another classic consequence of special relativity is immediately evident from the graphical representation of coordinates in Figure 14.9. Two events which occur at the same time in the unprimed frame, occur at different times in the primed frame. Imagine two events with the same value of $c_o t$, but different values of x. They would occur on a horizontal line in Figure 14.9. In the primed system, not only will they have different values of x', but they also will have different values of $c_o t'$.

The representation of Figure 14.8 for describing the transformations of special relativity does have problems, however. First, consider the magnitude of the basis vectors of the primed system. These covariant basis vectors are specified in terms of the unprimed basis vectors by Equation 14.120. Specifically for $\hat{g}_1'$:

$$\hat{g}_1' = \gamma_o \, \hat{g}_1 + \gamma_o \beta_o \, \hat{g}_2.$$ (14.123)

The magnitude squared of this basis vector is obtained by an inner product of the basis vector with itself:

$$
\begin{aligned}
\left|\hat{\mathbf{g}}_1'\right|^2 &= \hat{\mathbf{g}}_1' \cdot \hat{\mathbf{g}}_1'. \\
&= (\gamma_o \, \hat{\mathbf{g}}_1 + \gamma_o \beta_o \, \hat{\mathbf{g}}_2) \cdot (\gamma_o \, \hat{\mathbf{g}}_1 + \gamma_o \beta_o \, \hat{\mathbf{g}}_2) \\
&= \gamma_o^2 (1 + \beta_o^2),
\end{aligned}
\tag{14.124}
$$

where the last step follows because the unprimed basis vectors have unit magnitude. As long as the primed system is moving with respect to the unprimed system, that is if $v_o \neq 0$, Equation 14.124 says that $\hat{\mathbf{g}}_1'$ and, by a similar argument $\hat{\mathbf{g}}_2'$, are not unit vectors. This in itself is not that serious a problem, because our formalism does easily handle nonunit basis vectors. It does, however, complicate the geometrical picture of Figure 14.8, because you cannot simply axis-parallel-project a vector onto a basis vector to obtain a vector component. You must also divide by the magnitude of the corresponding basis vector.

A more serious problem arises when you consider the inner product in this system. In particular, let's look at $\bar{\mathbf{r}} \cdot \bar{\mathbf{r}}$, the inner product of the position vector with itself. This quantity is a scalar and should be invariant to transformations. In the unprimed frame,

$$
\bar{\mathbf{r}} = x \, \hat{\mathbf{g}}_1 + c_o t \, \hat{\mathbf{g}}_2,
\tag{14.125}
$$

so we expect the quantity $\bar{\mathbf{r}} \cdot \bar{\mathbf{r}}$ to be given by the expression

$$
x^2 + (c_o t)^2.
\tag{14.126}
$$

But we already determined in Equation 14.117 that special relativity requires the quantity $x^2 - (c_o t)^2$ to be invariant. It is not possible for both quantities to be invariant. The problem lies in our assumption about the geometry of the space. The value of $\bar{\mathbf{r}} \cdot \bar{\mathbf{r}}$ is not given by Equation 14.126! This is discussed in the next section.

The Metric of Special Relativity   Both the problem with the magnitudes of the primed covariant basis vectors and the invariance of the magnitude of the position vector involve some assumptions about how to take an inner product. In both cases, we took the inner product of a vector with itself in the unprimed coordinate system assuming an *Euclidian* metric of the form

$$
M_{ij} \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.
\tag{14.127}
$$

In effect, this was the metric used in the calculation of Equation 14.126:

$$
\begin{aligned}
\bar{\mathbf{r}} \cdot \bar{\mathbf{r}} &= \begin{bmatrix} x & c_o t \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ c_o t \end{bmatrix} \\
&= x^2 + (c_o t)^2.
\end{aligned}
\tag{14.128}
$$

With a little reflection, it can be seen that the invariant quantity of Equation 14.117 can be generated if we use an alternate metric:

$$M_{ij} \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{14.129}$$

When this metric is used in the unprimed system, Equation 14.126 is modified to

$$\bar{\mathbf{r}} \cdot \bar{\mathbf{r}} = \begin{bmatrix} x & c_o t \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ c_o t \end{bmatrix}$$

$$= x^2 - (c_o t)^2. \tag{14.130}$$

How does this new metric affect the orthogonality and the magnitudes of the basis vectors? Using the metric of Equation 14.129 to form the inner product of $\hat{\mathbf{g}}_1$ with itself gives

$$\hat{\mathbf{g}}_1 \cdot \hat{\mathbf{g}}_1 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$= 1. \tag{14.131}$$

The $\hat{\mathbf{g}}_1$ basis vector is still a unit vector. The inner product between $\hat{\mathbf{g}}_1$ and $\hat{\mathbf{g}}_2$ is

$$\hat{\mathbf{g}}_1 \cdot \hat{\mathbf{g}}_2 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= 0. \tag{14.132}$$

The covariant basis vectors of the unprimed frame are still orthogonal. But the magnitude of $\hat{\mathbf{g}}_2$ is

$$\hat{\mathbf{g}}_2 \cdot \hat{\mathbf{g}}_2 = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= -1. \tag{14.133}$$

This says that, while $|\hat{\mathbf{g}}_2| = 1$, $\hat{\mathbf{g}}_2$ itself must be imaginary! Very odd.

What about the basis vectors of the primed frame? Using the transformation equations given by Equation 14.120, we can write

$$\hat{\mathbf{g}}_1' = \gamma_o \, \hat{\mathbf{g}}_1 + \gamma_o \beta_o \, \hat{\mathbf{g}}_2 \tag{14.134}$$

$$\hat{\mathbf{g}}_2' = \gamma_o \beta_o \, \hat{\mathbf{g}}_1 + \gamma_o \, \hat{\mathbf{g}}_2. \tag{14.135}$$

Using the unprimed metric, the magnitude of $\hat{\mathbf{g}}_1'$ is determined by

$$\hat{\mathbf{g}}_1' \cdot \hat{\mathbf{g}}_1' = \begin{bmatrix} \gamma_o & \gamma_o \beta_o \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \gamma_o \\ \gamma_o \beta_o \end{bmatrix}$$

$$= 1. \tag{14.136}$$

So $\hat{\mathbf{g}}_1'$ has now become a unit vector. The inner product between $\hat{\mathbf{g}}_1'$ and $\hat{\mathbf{g}}_2'$ is

$$\hat{\mathbf{g}}_1' \cdot \hat{\mathbf{g}}_2' = \begin{bmatrix} \gamma_o & \gamma_o\beta_o \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \gamma_o\beta_o \\ \gamma_o \end{bmatrix}$$

$$= 0. \tag{14.137}$$

Now the primed, covariant basis vectors of the skewed system are also orthogonal! Finally, the magnitude of $\hat{\mathbf{g}}_2'$ is determined by

$$\hat{\mathbf{g}}_2' \cdot \hat{\mathbf{g}}_2' = \begin{bmatrix} \gamma_o\beta_o & \gamma_o \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \gamma_o\beta_o \\ \gamma_o \end{bmatrix}$$

$$= -1. \tag{14.138}$$

The $\hat{\mathbf{g}}_2'$ basis vector is an imaginary unit vector, similar to $\hat{\mathbf{g}}_2$.

The inner products in the primed system could have also been performed using M_{ij}', the metric of the primed system. Since the metric is a tensor, the primed metric is related to the unprimed metric in the normal manner:

$$M_{ij}' = g^m{}_i\, g^n{}_j\, M_{mn}. \tag{14.139}$$

In matrix notation, the primed metric is given by

$$[M'] = \begin{bmatrix} \gamma_o & \gamma_o\beta_o \\ \gamma_o\beta_o & \gamma_o \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \gamma_o & \gamma_o\beta_o \\ \gamma_o\beta_o & \gamma_o \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{14.140}$$

Notice the form of the metric is identical in the primed and unprimed systems. With this metric, the inner product between the covariant primed basis vectors can be determined:

$$\hat{\mathbf{g}}_1' \cdot \hat{\mathbf{g}}_1' = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$= 1, \tag{14.141}$$

$$\hat{\mathbf{g}}_1' \cdot \hat{\mathbf{g}}_2' = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= 0, \tag{14.142}$$

and

$$\hat{\mathbf{g}}_2' \cdot \hat{\mathbf{g}}_2' = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= -1. \tag{14.143}$$

These are exactly the same results obtained previously, using the unprimed metric.

The fact that one of the basis vectors is imaginary seems like a very strange thing. But there is nothing mathematically inconsistent with it. The distance along the the axis is now simply measured in terms of "imaginary" meter sticks instead of "real" ones. The new metric, while not very intuitive, solves our problems with the simple picture in Figure 14.9. The basis vectors are now unit vectors and the invariance of $x^2 - (c_o t)^2$ is built into the inner product operation. This weird geometry is often called Minkowski space, while the new metric is called the Minkowski metric.

We actually can devise an equivalent coordinate system which does not use the imaginary time axes. In this construction, we write the position vector as

$$\bar{\mathbf{r}} = x\,\hat{\mathbf{e}}_1 + ic_o t\,\hat{\mathbf{e}}_2, \tag{14.144}$$

where the basis vectors $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$ are now real, orthonormal, unit vectors. The same is true in the primed system. In this case inner products are performed with the "normal" Euclidian metric. This is equivalent to our earlier construction, using the imaginary basis vector, since all we have really done here is pull out a factor of i from the basis vector and put it into the coordinate. The problem with this view is that you lose the convenient representation of Figure 14.9. There is no way to interlace the two sets of axes, represent the same event in both systems as a single point, and visualize the transformation with axis-parallel projections. The best we can do is draw two distinct coordinate systems, as shown in Figure 14.11.

Extension to Three Dimensions   These ideas can easily be extended to three spatial dimensions. The position vector, in the unprimed and primed systems, becomes

$$\begin{aligned}
\bar{\mathbf{r}} &= x\,\hat{\mathbf{g}}_1 + y\,\hat{\mathbf{g}}_2 + z\,\hat{\mathbf{g}}_3 + c_o t\,\hat{\mathbf{g}}_4 \\
&= x'\,\hat{\mathbf{g}}_1' + y'\,\hat{\mathbf{g}}_2' + z'\,\hat{\mathbf{g}}_3' + c_o t'\,\hat{\mathbf{g}}_4'.
\end{aligned} \tag{14.145}$$

The Minkowski metric expands to

$$[M] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{14.146}$$
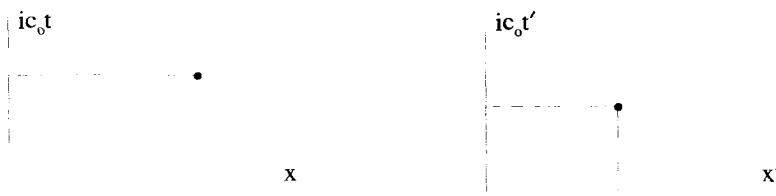
Figure 14.11   A Second Picture of the Coordinate Systems for Special Relativity

The Lorentz transformation between the two systems becomes

$$t^i{}_j = \begin{bmatrix} \gamma_o & 0 & 0 & -\gamma_o\beta_o \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma_o\beta_o & 0 & 0 & \gamma_o \end{bmatrix}. \tag{14.147}$$

In this expression, the relative motion between the coordinate systems is along the x direction. The appropriate transformation for an arbitrary direction of relative motion can be obtained by a rotational transformation of this matrix.

The position vector, as written in Equation 14.145, is usually called the position "four-vector." There are many other important four-vectors, some of which are developed in the exercises at the end of this chapter. One of their more useful properties is that the inner product between any two four-vectors is invariant to coordinate transformation.

EXERCISES FOR CHAPTER 14

1. Consider the primed and unprimed coordinate systems shown below:
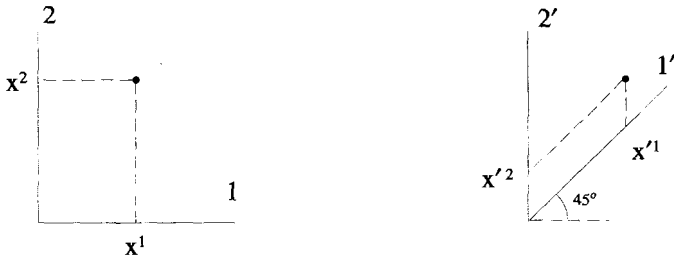
Let the two systems have a common origin. Each system has a set of covariant basis vectors $\hat{\mathbf{g}}_1$, $\hat{\mathbf{g}}_2$ and $\hat{\mathbf{g}}'_1$, $\hat{\mathbf{g}}'_2$.

(a) Find the expressions for x^1 and x^2 in terms of x'^1 and x'^2 and the inverse relations.

(b) Identify the covariant basis vectors in the primed system by drawing their magnitudes and directions.

(c) The vector $\overline{\mathbf{V}}$, expressed in the unprimed system, is given by

$$\overline{\mathbf{V}} = 2\hat{\mathbf{g}}_1 + \hat{\mathbf{g}}_2.$$

What are the contravariant and covariant components of this vector in the primed system?

(d) Using its contravariant and covariant components, determine the magnitude of the vector $\overline{\mathbf{V}}$ by forming the inner product of the vector with itself. Show

that the same result is obtained if just the contravariant components are used
with the appropriate metric.

2. Consider the unprimed, two-dimensional Cartesian system and a primed, skewed
   system, as shown in the figure below:
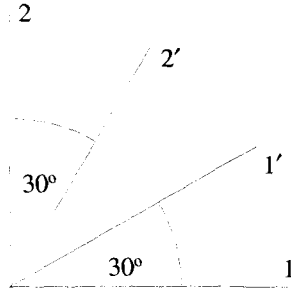
   (a) Determine the set of equations that relate the unprimed coordinates to the
       primed coordinates.
   (b) Determine the elements of the transformation matrices $t^i{}_j$ and $g^i{}_j$.
   (c) Determine the elements of the M'_{ij} metric in the primed system, using the
       elements of the transformation matrix $g^i{}_j$ and Equation 14.92.
   (d) Show that the elements of the metric in part (c) could also have been obtained
       by using $M'_{ij} = \hat{\mathbf{g}}'_i \cdot \hat{\mathbf{g}}'_j$.

3. We have shown that the metric is a tensor. It can be expressed using purely
   contravariant components as M^{ij}, or with purely covariant components as M_{ij},
   so that

   $$\overline{\overline{\mathbf{M}}} = M_{ij}\hat{\mathbf{g}}^1\hat{\mathbf{g}}^2 = M^{ij}\hat{\mathbf{g}}_1\hat{\mathbf{g}}_2.$$

   (a) Evaluate the product $M^{ik}M_{kj}$.
   (b) Use the result for part (a) to simplify the following expressions:
       i. $V_i M^{ij} M_{jk}$.
       ii. $V^k M^{ij} M_{jk}$.

4. Show that the cross product $\overline{\mathbf{A}} \times \overline{\mathbf{B}}$ expressed in a skewed system can be written
   as

   $$A^i B^j \hat{\mathbf{g}}^k \nu\epsilon_{ijk}$$

   and determine ν.

5. Starting with Equation 14.92 and taking the unprimed system to be orthonormal,
   show that

   $$M'_{ij} = \left(\frac{\partial x^m}{\partial x'^i}\right)\left(\frac{\partial x^m}{\partial x'^j}\right).$$

The definition of the metric for the primed system is $M'_{ij} = \hat{\mathbf{g}}'_i \cdot \hat{\mathbf{g}}'_j$, which depends only on the properties of the primed system. The above equation, however, seem to also imply that the primed metric depends on the unprimed system. Explain this apparent contradiction.

6. In this problem, the ideas of contravariance and covariance will be used to analyze a two-dimensional polar system. Designate a two-dimensional, Cartesian system with coordinates $(x, y) \rightarrow (x^1, x^2)$ as the unprimed system. Let the polar system be primed, with coordinates $(\rho, \theta) \rightarrow (x'^1, x'^2)$. The relationships between the two sets of coordinates are

$$x'^1 = \sqrt{(x^1)^2 + (x^2)^2}$$

$$x'^2 = \tan^{-1}\left[\frac{x^2}{x^1}\right].$$

(a) Obtain the elements of the metric for the primed polar system, M'_{ij} using the expression $M'_{ij} = g^l{}_i g^l{}_j$ where $g^l{}_i = \partial x^l / \partial x'^i$.

(b) Show for the metric you obtained in part (b) that

$$M'_{ij} \neq \hat{\mathbf{e}}'_i \cdot \hat{\mathbf{e}}'_j,$$

where the $\hat{\mathbf{e}}'_i$ are the traditional polar basis vectors $(\hat{\mathbf{e}}_\rho, \hat{\mathbf{e}}_\theta)$ defined in Chapter 3.

(c) Using contravariant/covariant transformation techniques, obtain the contravariant $(\hat{\mathbf{g}}^\rho, \hat{\mathbf{g}}^\theta)$ and covariant $(\hat{\mathbf{g}}_\rho, \hat{\mathbf{g}}_\theta)$ basis vectors of the polar system in terms of the Cartesian basis vectors $(\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y)$. Compare these contravariant and covariant polar basis vectors to the traditional polar basis vectors $(\hat{\mathbf{e}}_\rho, \hat{\mathbf{e}}_\theta)$. Pick a point in the first quadrant and plot $\hat{\mathbf{e}}_x$, $\hat{\mathbf{e}}_y$, $\hat{\mathbf{e}}_\rho$, $\hat{\mathbf{e}}_\theta$, $\hat{\mathbf{g}}_\rho$, and $\hat{\mathbf{g}}_\theta$.

(d) The displacement vector can be expressed using the covariant polar basis vectors as

$$d\bar{\mathbf{r}} = dx^k \hat{\mathbf{g}}'_i = d\rho\hat{\mathbf{g}}_\rho + d\theta\hat{\mathbf{g}}_\theta,$$

so that the contravariant components of the displacement vector in the polar system are simply $d\rho$ and $d\theta$. What are the covariant components of the displacement vector in the polar system?

(e) Using the contravariant and covariant components of the displacement vector in the polar system, form the inner product $d\bar{\mathbf{r}} \cdot d\bar{\mathbf{r}}$.

7. Show that if a set of primed coordinates (x, ct) are related to a set of unprimed coordinates (x', ct') by a Lorentz transformation, then

$$x^2 - (ct)^2 = x'^2 - (ct')^2.$$

8. This problem illustrates the differences between Euclidian geometry and Minkowski geometry. First for the Minkowski geometry:

(a) Draw an orthonormal set of axes for an (x, ct)-system and plot the point $x = 2cm$, $ct = 3cm$.

(b) On the same plot, draw the x' and ct' axes of a primed system, moving at velocity $v = .58c$ with respect to the unprimed system.

(c) Transform the coordinates of the point in part (a) into the primed system and indicate the values of (x', ct') on the axes of your drawing.

Now for Euclidian geometry:

(a) Start with a (x^1, x^2)-system that is orthonormal. Make a labeled sketch of the axes of this system and plot the point $x^1 = 2cm$, $x^2 = 3cm$.

(b) Determine the $g^i{}_j$ and the $t^i{}_j$ transformation matrices that relate the $(x^1, x^2)-$ system to a skewed primed system, where the x'^1 and x'^2 axes lie in the same position as the x' and ct' axes of the above Minkowski system.

(c) Transform the point in part (a) into the primed coordinate system and indicate the values of (x'^1, x'^2) on the axes of your drawing.

(d) Notice and comment on the differences in scaling of the x' and ct' axes of the Minkowski system and the scaling of the primed axes of the Euclidian system.

9. Consider a wave propagating along the x-axis with the a space-time dependence of

$$e^{i(kx + \omega t)}.$$

The time period of the wave is $T = 2\pi/\omega$, and the spatial wavelength is $\lambda = 2\pi/k$.

(a) Assume that ω and k are positive real numbers. Is the wave traveling in the positive or negative x-direction?

(b) Now observe this wave from a coordinate system moving along the x-direction, with a positive velocity $v = \beta c$ with respect to the original system. Should the frequency of the wave in the moving system be larger or smaller than ω?

(c) Call the moving frame the primed frame. In this system, the space-time dependence of the wave becomes

$$e^{i(k'x' + \omega't')}.$$

The primed quantities can be obtained from the unprimed quantities by using the transformation techniques of special relativity. First, it is necessary to construct a pair of "four-vectors" (actually, since we have only one spatial dimension in this problem, they are "two-vectors"). The terms in the exponential brackets can be expressed as the inner product between two of these vectors, $\overline{\mathbf{R}}$ and $\overline{\mathbf{K}}$. $\overline{\mathbf{R}}$ is the position two-vector,

$$\overline{\mathbf{R}} = x\hat{\mathbf{g}}_1 + ct\hat{\mathbf{g}}_2,$$

while $\overline{\mathbf{K}}$ is a frequency two-vector with the components

$$\overline{\mathbf{K}} = k\hat{\mathbf{g}}^1 + (\omega/c)\hat{\mathbf{g}}^2.$$

Why are these k and ω/c components covariant quantities?

(d) With these two-vectors defined, their inner product is

$$\overline{\mathbf{K}} \cdot \overline{\mathbf{R}} = kx + \omega t,$$

which is invariant to transformation. Since we are working in Minkowski space, why wasn't the metric used to form this inner product?

(e) The invariance of the quantity in part (d) says that

$$kx + \omega t = k'x' + \omega't'.$$

What is the physical interpretation of this invariance?

(f) Use the Lorentz transformations to obtain expressions for k' and ω'/c in terms of k and ω/c and identify the relativistic Doppler shift of the frequency and the Lorentz contraction for the wavelength as observed in the moving frame. Does this agree with your intuitive answer to part (b)?

(g) Show that if $\beta \ll 1$ your answer to part (f) goes to the classical limit

$$\omega' = \omega + kv$$
$$k' = k.$$

10. Why does Equation 14.122 imply that you cannot accelerate an object past the speed of light?

15

INTRODUCTION TO GROUP THEORY

Group theory provides a formal framework for taking advantage of the different types of symmetries that pervade science. In this chapter, we introduce the mathematical definition of a group, and present several examples of both discrete and continuous groups that occur in physics and chemistry.

15.1 THE DEFINITION OF A GROUP

A group is a set of abstract elements which can be "multiplied" together. In this context, multiplication does not necessarily refer to the common operation between real or complex numbers, but rather a more generalized operation which combines two elements of the group together. To determine if something is actually a group, you must first know both the group elements and the rule for multiplication.

We will typically denote an entire group using a capital letter, while the distinct elements that make up the group are written with a small letter and a subscript. For example, the group G might have the elements $g_1, g_2, g_3, \ldots, g_h$. To talk about a single, but arbitrary group element, we will often use a letter subscript. For example, g_i refers to any one of the elements $g_1, g_2, g_3, \ldots$

In order for G to be called a group, its elements must obey the following four rules:

1. If g_i and g_j are elements of G, then their product, $g_k = g_i \, g_j$, is also an element of G.
2. The multiplication operation is associative: $(g_i \, g_j)g_k = g_i(g_j \, g_k)$.
3. One element I serves as an identity element, such that $I \, g_i = g_i \, I = g_i$ for all g_i in G.

4. Each element of G has an inverse, $g_j = g_i^{-1}$, that is also an element of G. The product of g_i and its inverse equals the identity element: $g_i g_i^{-1} = g_i^{-1} g_i = I$.

Keep in mind that rule #2 does not require that the multiplication be commutative, although it may be. That is, $g_i g_j$ does not have to equal $g_j g_i$. If the multiplication is commutative for all combinations of group elements, the group is called *Abelian*.

Where possible, we will identify the first group element as the identity element, whose existence is required by rule #3. That is, $g_1 \equiv I$. Notice all operations with this particular element are necessarily commutative.

15.2 FINITE GROUPS AND THEIR REPRESENTATIONS

In this section we present several examples of groups. As we progress, we will gradually accumulate an arsenal of terminology to keep track of the different types of groups and their properties.

15.2.1 The Cyclic Group C_4

Our first example of a group consists of the elements

$$G = \{I, \quad g_2, \quad g_3, \quad g_4\}. \tag{15.1}$$

The *order* of a group, which we will call h, is the number of distinct elements it contains. The group described in Equation 15.1 has $h = 4$. If h is finite, as it is in this case, the elements form a *finite group* and the elements are discrete. In contrast, *infinite groups* have $h \to \infty$. Infinite groups can have discrete or continuous elements. In this chapter, we will limit our discussions to discrete, finite groups and continuous, infinite groups.

The multiplication properties of a finite group can be summarized concisely using a *multiplication table*. The row label of a table identifies the first element in the product, while the column label represents the second. If we always make the first group element the identity, it is trivial to construct the first row and column of any table, as shown for our four-element group in Table 15.1. The first row and column of all multiplication tables, for groups of the same order, will be identical. The rest of the table depends upon the particular group. For the C_4 group, the multiplication table is filled in with a cyclic arrangement of the group elements, as shown in Table 15.2.

TABLE 15.1. Start of a Multiplication Table

	I	g_2	g_3	g_4
I	I	g_2	g_3	g_4
g_2	g_2			
g_3	g_3			
g_4	g_4			

TABLE 15.2. Multiplication Table for C_4

C_4	I	g_2	g_3	g_4
I	I	g_2	g_3	g_4
g_2	g_2	g_3	g_4	I
g_3	g_3	g_4	I	g_2
g_4	g_4	I	g_2	g_3

Notice this table is symmetric about the diagonal and therefore is Abelian. A good exercise would be to convince yourself that this table satisfies all the requirements of a group.

Notice that each group element appears exactly once in each row and column of Table 15.2. This is not a coincidence, but rather a fundamental property of all group multiplication tables. To see this, assume the products of a group element g_i with two different group elements, g_j and g_k, produce the same element g_m. In other words,

$$g_i\, g_j = g_m \tag{15.2}$$

$$g_i\, g_k = g_m. \tag{15.3}$$

If this were the case, g_m would appear more than once in the ith row of the table. If we multiply both these equations by g_i^{-1}, we get the pair of equations

$$g_j = g_i^{-1}\, g_m \tag{15.4}$$

$$g_k = g_i^{-1}\, g_m, \tag{15.5}$$

which imply

$$g_j = g_k. \tag{15.6}$$

This clearly violates our initial assumption that $g_j \neq g_k$. This proof shows that an element cannot appear twice in a single row. It is trivial to construct a similar proof for columns. The combination of both results is often called the *rearrangement lemma*, because each row and column of a table must contain all the group elements, but with different ordering.

Figure 15.1 shows a simple two-dimensional picture that can be associated with the C_4 group. Each group element is a rigid rotation of the object in the plane of the figure, which leaves the square array in the same physical configuration it was initially. These operations are sometimes called symmetry operations. A rigid rotation requires the relative distances between the individual objects that make up the body remain constant. In the case of the square object of Figure 15.1, imagine sticks between the four objects in the corners of the square. The I element clearly does nothing. The other elements are integer multiples of $\pi/2$ rotations of the figure around its center. The group multiplication table shows how two successive symmetry operations are equivalent to another symmetry operation. For example, two $\pi/2$ rotations are
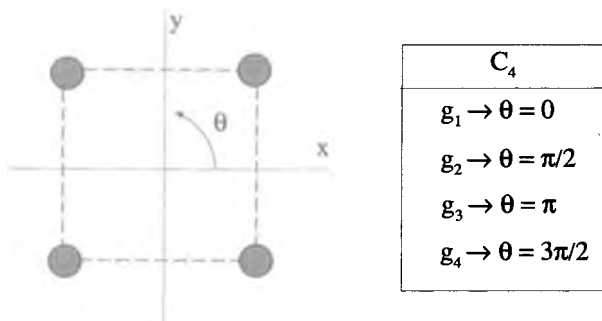
Figure 15.1 A Graphical Interpretation of the C_4 Elements

equivalent to a single π rotation. With this picture, the origin of the term *cyclic* should be a little more obvious.

A *representation* of a group is a mapping between the group elements and a particular set of quantities, which obey the same multiplication table. The most useful type of representations use matrices as the group elements, with the multiplication between elements accomplished using standard matrix multiplication rules. Any finite group has an infinite number of different matrix representations, but as we will see later in the chapter, a finite number of "important" ones.

Our picture of the rotations of the square array in Figure 15.1 immediately leads to a simple 1×1 representation of the C_4 group. We can associate four complex phasors with the C_4 elements, as shown in Table 15.3. Using this representation for the elements of C_4, the multiplication table fills in as shown in Table 15.4. Notice how this table has exactly the same structure as Table 15.2. If we wished, we could even drop the $e^{i\theta}$ notation and write the elements of this representation as shown in Table 15.5. This does not change the multiplication table in any way.

A different representation for the C_4 group can also be devised by using Figure 15.1 as a guide for constructing a set of two-dimensional transformation matrices.

TABLE 15.3. A Representation for C_4

I	g_2	g_3	g_4
e^{i0}	$e^{i\pi/2}$	$e^{i\pi}$	$e^{i3\pi/2}$

TABLE 15.4. C_4 Representation Satisfying its Multiplication Table

	e^{i0}	$e^{i\pi/2}$	$e^{i\pi}$	$e^{i3\pi/2}$
e^{i0}	e^{i0}	$e^{i\pi/2}$	$e^{i\pi}$	$e^{i3\pi/2}$
$e^{i\pi/2}$	$e^{i\pi/2}$	$e^{i\pi}$	$e^{i3\pi/2}$	e^{i0}
$e^{i\pi}$	$e^{i\pi}$	$e^{i3\pi/2}$	e^{i0}	$e^{i\pi/2}$
$e^{i3\pi/2}$	$e^{i3\pi/2}$	e^{i0}	$e^{i\pi/2}$	$e^{i\pi}$

TABLE 15.5. Another Form of the C_4 Representation

I	g_2	g_3	g_4
1	i	-1	$-i$
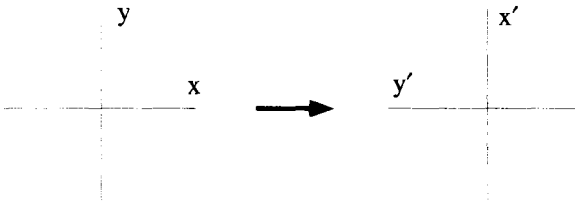
Figure 15.2 $\pi/2$ Rotation for g_2 of the C_4 Group

Since the g_2 element corresponds to a counterclockwise rotation of $\pi/2$, its effect on a two-dimensional orthonormal coordinate system is shown in Figure 15.2. The 2×2 transformation matrix that relates the primed coordinates to the unprimed coordinates for this rotation is simply

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{15.7}$$

By associating the other group elements with their transformation matrices in a similar manner, we can generate another valid representation of C_4. The elements of this 2×2 matrix representation for C_4 are shown in Table 15.6. It is easy to check that the original multiplication table is still satisfied by multiplying the matrices together.

It is clear from Figure 15.1, that we can define a cyclic group C_h of arbitrary order h, by increasing the number of sides of the polygon. The representations discussed here are also easily extended in the same manner.

15.2.2 The Vierergruppe D_2

The only other possible group of order four has the multiplication table shown in Table 15.7. This group is called the vierergruppe and does not demonstrate the cyclic nature of C_4. Instead, each element is its own inverse, and the product of any two

TABLE 15.6. A Different Representation for C_4

I	g_2	g_3	g_4
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

TABLE 15.7. The Vierergruppe Multiplication Table

D_2	I	g_2	g_3	g_4
I	I	g_2	g_3	g_4
g_2	g_2	I	g_4	g_3
g_3	g_3	g_4	I	g_2
g_4	g_4	g_3	g_2	I

of the elements, excluding the identity, generates the third. The symmetry about the diagonal still exists, so the vierergruppe is Abelian.

As before, a simple physical picture can be associated with this group. Imagine an object, perhaps a molecule, made of three pairs of different atoms. The six atoms are arranged as shown in Figure 15.3. Each element of the vierergruppe can be associated with a rigid rotation of the system which leaves it in an identical state. Clearly, rotations of π around any of the axes are group elements. The fourth element is, of course, the identity. We write these elements as $R(0)$, $R_x(\pi)$, $R_y(\pi)$, and $R_z(\pi)$. We will often make use of this "R" notation for rotations. The subscript refers to the axis of rotation, while the argument is the angle of rotation, counterclockwise about that axis. The correspondence between the vierergruppe multiplication table and this picture needs to be shown. Since applying a π rotation twice is the same as no rotation at all, the fact that each element is its own inverse is obvious. That the product of any two nonidentity elements generates the third is demonstrated by Figure 15.4 for the $R_x(\pi) R_y(\pi) = R_z(\pi)$ operation. The other permutations follow in an analogous manner. Because of this picture, the vierergruppe is often called D_2. The D stands for dihedral, a figure formed by the intersection of two planes, and the 2 indicates there is a twofold symmetry around each axis.

We can easily construct a 3×3 matrix representation of D_2 by considering the effect of the symmetry operations on the coordinate axes shown in Figure 15.3.
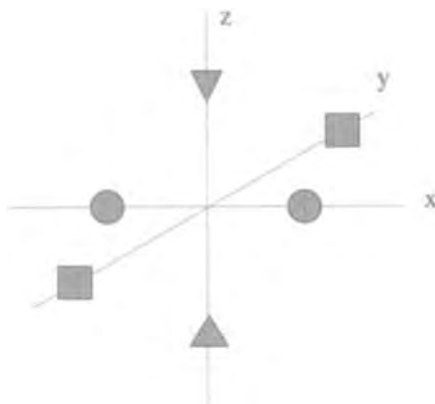
Figure 15.3   Symmetry for the Vierergruppe
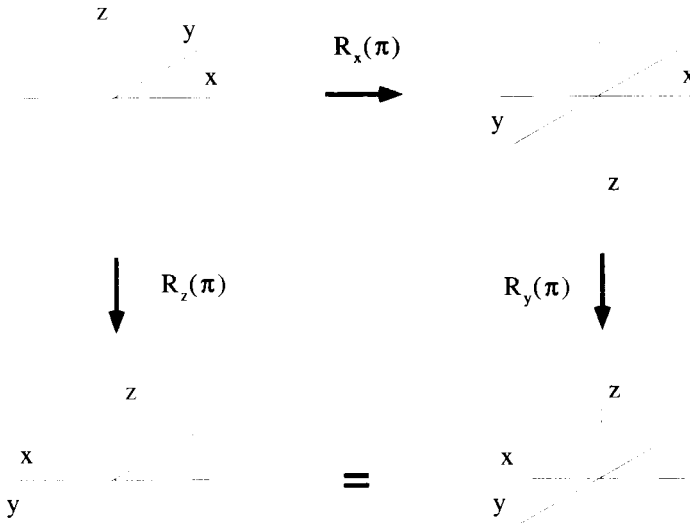
Figure 15.4   Demonstration of Vierergruppe Element Multiplication

For example, consider the group element $R_x(\pi)$. Figure 15.5 shows its effect on the coordinate system. The matrix that converts the unprimed coordinates to the primed coordinates in this picture is

$$
\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} .
\tag{15.8}
$$

The matrices for the remaining elements can be generated in a similar manner, and the complete representation is given in Table 15.8.
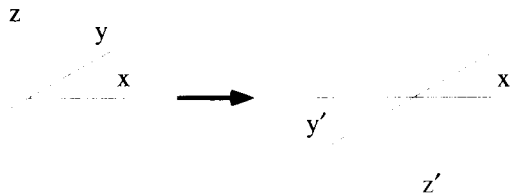
Figure 15.5   Vierergruppe Symmetry Rotation About the x-Axis

TABLE 15.8. A 3 × 3 Representation for the Vierergruppe

I	g_2	g_3	g_4
$R(0)$	$R_x(\pi)$	$R_y(\pi)$	$R_z(\pi)$
$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
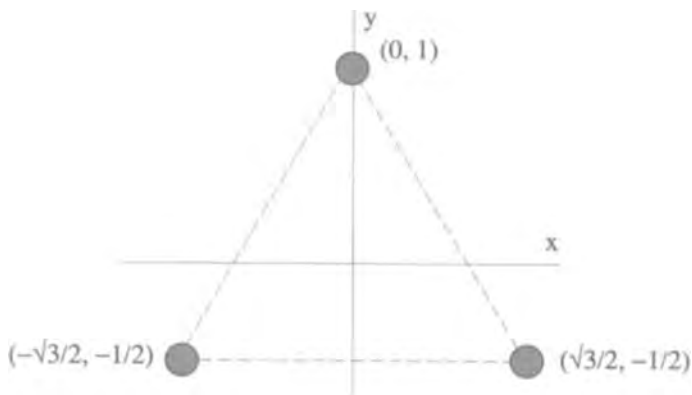
Figure 15.6   Object for Threefold Symmetry

15.2.3   The Threefold Symmetry Group D_3

Figure 15.6 shows an object with a threefold symmetry. Three identical objects are placed at the corners of an equilateral triangle situated in the xy-plane. The coordinates of each point are labeled in this figure.

What are the rigid rotations that leave this object unchanged? If we restrict ourselves to rotations in the xy-plane, the transformations constitute the C_3 group, a version of the cyclic group we investigated earlier. Using the "R" notation, the elements are $R_z(0)$, $R_z(2\pi/3)$, and $R_z(4\pi/3)$. We can write down a simple matrix representation of these elements, by determining the 2×2 transformation matrices associated with each rotation, the C_3 version of the representation given in Table 15.6. The result is given in Table 15.9.

TABLE 15.9. The Three Elements of C_3 and a Representation

I	g_2	g_3
$R(0)$	$R_z(2\pi/3)$	$R_z(4\pi/3)$
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{bmatrix}$	$\begin{bmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{bmatrix}$
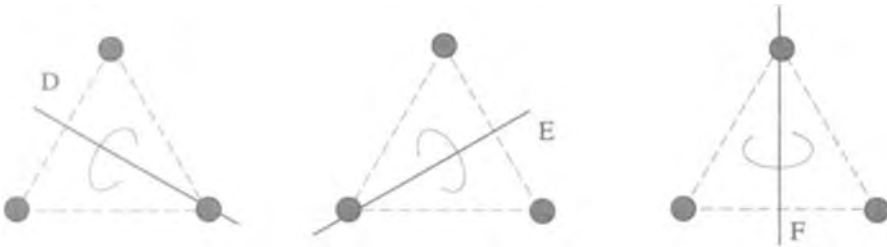
Figure 15.7 Axes for Interchanging Two of the Threefold Symmetric Objects

TABLE 15.10. The Six Group Elements of a D_3 Representation

I	g_2	g_3	g_4	g_5	g_6
$R(0)$	$R_z(2\pi/3)$	$R_z(4\pi/3)$	$R_D(\pi)$	$R_E(\pi)$	$R_F(\pi)$
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\frac{1}{2}\begin{bmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}$	$\frac{1}{2}\begin{bmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{bmatrix}$	$\frac{1}{2}\begin{bmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & -1 \end{bmatrix}$	$\frac{1}{2}\begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$

There are other rigid body transformations that leave the configuration of this object unchanged. If rotations in three dimensions are allowed, there are three more symmetry operations that interchange only two vertices of the triangle. These can be viewed as rotations of π about the D, E, and F axes, shown in Figure 15.7. Now a total of six group elements have been generated, which are listed along with their matrix representation in Table 15.10.

To finish the story, we need to confirm that these elements form a complete group, and determine the multiplication table. This can be done using pictures similar to Figure 15.4 or by multiplying each possible pair of the matrix representations together. With either method, the result is that these six elements do indeed form a complete group, and the multiplication table is given in Table 15.11. This group is called D_3. Notice that it is not Abelian.

TABLE 15.11. The Multiplication Table for D_3

D_3	I	g_2	g_3	g_4	g_5	g_6
I	I	g_2	g_3	g_4	g_5	g_6
g_2	g_2	g_3	I	g_5	g_6	g_4
g_3	g_3	I	g_2	g_6	g_4	g_5
g_4	g_4	g_6	g_5	I	g_3	g_2
g_5	g_5	g_4	g_6	g_2	I	g_3
g_6	g_6	g_5	g_4	g_3	g_2	I
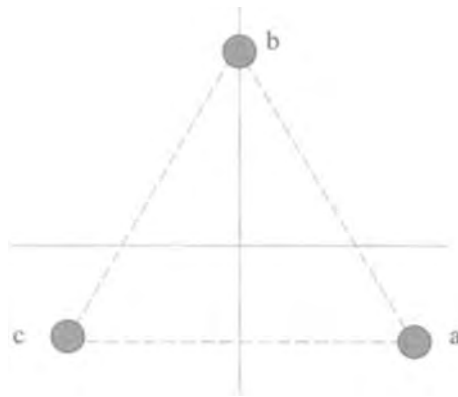
Figure 15.8 Labeling for Operator Representation of the D_3 Group Elements

15.2.4 Operator Representation and Permutation Groups

Another useful type of group representation has operators as the group elements. As an example of this type of representation, consider the D_3 group of the previous section. To develop the operator elements of this group we label the objects at the corners of the triangle with the letters a, b, and c, as shown in Figure 15.8. The g_2 operation takes the sequence (abc) to the sequence (cab), while the g_4 element changes the sequence (abc) to (acb). We can define a "rearrangement" operator [132] such that

$$[132](abc) \rightarrow (acb). \tag{15.9}$$

What this notation means is that the first value between the initial parentheses gets mapped onto the first value between the final parentheses. The second and third values are interchanged. All of the elements of the D_3 group can be expressed as a set of these rearrangement operators as described by Table 15.12.

Multiplications for the elements of D_3 are easily performed with these operators. The product of the two elements

$$g_2\, g_4 = g_5 \tag{15.10}$$

TABLE 15.12. Operator Representation for D_3

I		g_2		g_3		g_4		g_5		g_6	
b		a		c		c		a		b	
•		•		•		•		•		•	
•	•	•	•	•	•	•	•	•	•	•	•
c	a	b	c	a	b	b	a	c	b	a	c
[123]		[312]		[231]		[132]		[213]		[321]	

is simply written as

$$[312][132] = [213]. \qquad (15.11)$$

Notice how the product of two elements is now effectively built into the notation. We do not need to reference a multiplication table to determine the result given in Equation 15.11, because the first operator [312] applied directly to [132] gives the result [213].

The operator representation is very useful for discussing the general class of *permutation groups*. The S_n permutation group describes the indistinguishable ways that n identical objects can be arranged. The identity element of this group can be represented by the operator $[1234 \cdots n]$, while the other elements have operators that are generated by all the possible permutations of the numbers 1 through n. The S_n group is therefore of order $n!$.

15.3 SUBGROUPS, COSETS, CLASS, AND CHARACTER

There are several ways of sorting the elements of a group into categories. This section describes the four most important classifications for group theory: subgroups, cosets, class, and character.

15.3.1 Subgroups

A subgroup is a subset of the elements of a group that, by themselves, form a legitimate group. The D_3 group, whose multiplication table was given in Table 15.11, has several subgroups. One, consisting of the elements I, g_2, and g_3, is easy to spot in the upper left-hand corner of the table, and its multiplication table is shown in Table 15.13. There are three other subgroups that consist of two elements each; a subgroup formed by I and g_4, one from I and g_5, and the last from I and g_6. Each of these two element subgroups satisfies a multiplication table similar to the one shown in Table 15.14. Notice the identity element must be a part of any subgroup. Actually, if we

TABLE 15.13. Multiplication Table for a Subgroup of D_3

	I	g_2	g_3
I	I	g_2	g_3
g_2	g_2	g_3	I
g_3	g_3	I	g_2

TABLE 15.14. Another Multiplication Table for a Subgroup of D_3

	I	g_4
I	I	g_4
g_4	g_4	I

want to be rigorous, D_3 has two additional "trivial" subgroups. First, the entire group itself is considered a subgroup. Also, the identity element all by itself is a subgroup. These trivial cases occur for all groups, and are usually ignored.

15.3.2   Cosets

A coset is the set of elements formed by multiplying all the elements of a subgroup by an element of the group. Let the group G have order h and consist of the elements $g_1, g_2, \cdots g_h$. For the purposes of this discussion we will use the symbol G to represent all the elements of the group, in any sequence:

$$G = \{g_1, \ g_2, \ g_3, \ \cdots \ g_h\} = \{g_3, \ g_h, \ g_1, \ \cdots \ g_2\} = \text{etc.} \qquad (15.12)$$

This convention lets us write

$$g_i \, G = G, \qquad (15.13)$$

which is just a compact way of writing the rearrangement lemma. Assume G has a subgroup S of order h', with elements

$$S = \{s_1, \ s_2, \ s_3, \ \cdots \ s_{h'}\}. \qquad (15.14)$$

Let g_x be an element of G that may or may not be in S. If we premultiply g_x by all the different members of S, we get a set of h' distinct elements called a *right coset* of S. We write the collection of products as $S \, g_x$:

$$S \, g_x = \{s_1 \, g_x, \ s_2 \, g_x, \ s_3 \, g_x, \ \cdots \ s_{h'} \, g_x\}. \qquad (15.15)$$

Notice that this set of elements does not necessarily form a group. If g_x is in S, then the coset is equivalent to the group S. If g_x is not in S, every element of the coset must be an element that is not in S, and the coset cannot form a group because the identity element is missing. There are h possible choices for g_x, so there are h possible right cosets of S. If instead, we put g_x first in the product operation, we get a *left coset*:

$$g_x \, S = \{g_x \, s_1, \ g_x \, s_2, \ g_x \, s_3, \ \cdots \ g_x \, s_{h'}\}. \qquad (15.16)$$

Again, there are h possible left cosets of S.

It should be noted that some texts define cosets with the additional requirement that g_x not be an element of S. The distinction is not very important, since it only eliminates the cosets that are equivalent to S. Our choice for forming the cosets helps to clarify the following discussion.

The right and left cosets of S are not necessarily distinct. An important theorem says that the elements of two cosets of S either have all the same elements, or no elements in common. To see this, consider two right cosets, Sg_x and Sg_y. Now imagine these two cosets have an element in common. That is, for some choice of s_i and s_j,

$$s_i \, g_x = s_j \, g_y. \qquad (15.17)$$

Premultiplying this expression by s_j^{-1} and postmultiplying by g_x^{-1} gives

$$s_j^{-1} s_i = g_y g_x^{-1}. \qquad (15.18)$$

The element generated on the LHS of this equation is an element of S, so the product $g_y g_x^{-1}$ on the RHS must be an expression for this same element of S. Thus we can write

$$S g_y g_x^{-1} = S, \qquad (15.19)$$

or equivalently

$$S g_y = S g_x. \qquad (15.20)$$

Consequently, if the same element appears in two different cosets, the two cosets have all their elements in common. Conversely, if a particular element is present in one coset but not another coset, there can be no common elements between these two cosets.

The elements in any single coset are distinct. That is, no element can appear more than once in any given coset. This is a direct result of the rearrangement lemma—every element must appear once and only once in every row and column of the table. Therefore, every coset contains exactly h' different group elements. Therefore, because any two cosets will either have all elements in common or no elements in common, the order of the group h must be an integer multiple of the number of elements in any of its subgroups. That is,

$$h = nh' \quad \text{where } n = \text{a positive integer.} \qquad (15.21)$$

This is demonstrated quite nicely by the subgroups of D_3, discussed earlier. The D_3 group has six elements, while its nontrivial subgroups have orders two and three.

Example 15.1   As an example, consider the right cosets formed from one of the subgroups of D_3. Remember D_3 consists of six elements, which obey the multiplication Table 15.11. Earlier we found a subgroup formed from the first three elements of this main group:

$$S = \{I, g_2, g_3\}. \qquad (15.22)$$

There are six right cosets associated with this subgroup, each with three elements, as shown in Table 15.15. The first three cosets are formed by multiplying the subgroup

TABLE 15.15.   Right Cosets of a Subgroup of D_3

$S\ g_1$	$S\ g_2$	$S\ g_3$	$S\ g_4$	$S\ g_5$	$S\ g_6$
$I\ I = I$	$I\ g_2 = g_2$	$I\ g_3 = g_3$	$I\ g_4 = g_4$	$I\ g_5 = g_5$	$I\ g_6 = g_6$
$g_2\ I = g_2$	$g_2\ g_2 = g_3$	$g_2\ g_3 = I$	$g_2\ g_4 = g_5$	$g_2\ g_5 = g_6$	$g_2\ g_6 = g_4$
$g_3\ I = g_3$	$g_3\ g_2 = I$	$g_3\ g_3 = g_2$	$g_3\ g_4 = g_6$	$g_3\ g_5 = g_4$	$g_3\ g_6 = g_5$

by an element in the subgroup. These cosets therefore always contain the same three elements as the subgroup S. The last three cosets are formed by products of S with group elements that are not in S. Each of these cosets contain three distinct elements of G that are not in S.

15.3.3   Class and Character

Another way to sort group elements is by their *class*. If a and b are elements of a group, they are in the same class if we can find another group element g such that

$$g^{-1} a g = b. \tag{15.23}$$

This type of process, which involves an operator and its inverse acting on a, is called a similarity transform of a. Notice that the identity element must always be in a class by itself, because any similarity transform of the identity element will always generate the identity element.

In matrix algebra, the trace of a square matrix is the sum of all its diagonal elements. In subscript notation, it can be written

$$Tr[b] = b_{ii}, \tag{15.24}$$

where the doubled subscript implies a sum over the index i. It is easy to prove the matrix representations of all the group elements in one class must have the same trace. Suppose $[a]$ and $[b]$ are two matrix elements of a group, related by a similarity transform using group matrix elements $[g]$ and $[g]^{-1}$, in the matrix form of Equation 15.23,

$$[g]^{-1}[a][g] = [b], \tag{15.25}$$

or in subscript notation:

$$g_{ij}^{-1} a_{jk} g_{kl} = b_{il}. \tag{15.26}$$

The trace of $[b]$ becomes

$$b_{ii} = g_{ij}^{-1} a_{jk} g_{ki} \tag{15.27}$$

$$= g_{ki} g_{ij}^{-1} a_{jk} \tag{15.28}$$

$$= \delta_{kj} a_{jk} \tag{15.29}$$

$$= a_{kk}. \tag{15.30}$$

Thus $[a]$ and $[b]$ must have the same trace if they are in the same class.

This property turns out to be so important that the trace of a matrix representing a group element is given its own special name. In group theory, it is called the *character* of the element, and in this text is denoted as χ. We have shown that matrix elements

representing group elements of the same class will always have the same character. Keep in mind this does not always work the other way around. It is possible to have two matrix elements that are in different classes, but have the same character.

Example 15.2   The group D_3 has three different classes. As always, the identity element is in a class by itself. The two elements g_2 and g_3 are in a second class, while the elements g_4, g_5, and g_6 form a third class. This can be easily confirmed by performing the similarity transformations. If you look back at Figure 15.6, it can be seen that these three different classes for D_3 correspond to three very different types of symmetry operations. The identity element is unique because it is the only operation that represents no change to the orientation. The elements g_2 and g_3 are both operations which rotate the system in Figure 15.6 in the plane of the page, interchanging all three of the vertices of the triangle. The elements g_4, g_5 and g_6 are operations where only two of the objects are interchanged, while a third is held fixed. This gives an intuitive feel for what a class really is. It is a collection of all the symmetry operations which do "the same type of thing."

A list of the characters for a 2×2 matrix representation for D_3 are shown in Table 15.16. The first column of this table contains the 2×2 matrix representation of the group elements we found in Table 15.10. The second column indicates the character

TABLE 15.16. List of Characters for a D_3
Representation

g	χ
$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	2
$g_2 = \frac{1}{2} \begin{bmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}$	-1
$g_3 = \frac{1}{2} \begin{bmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{bmatrix}$	-1
$g_4 = \frac{1}{2} \begin{bmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & -1 \end{bmatrix}$	0
$g_5 = \frac{1}{2} \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}$	0
$g_6 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$	0

of each matrix element. Notice how we have subdivided the elements by their class. For this representation, each class has a different character, but remember this will not always be the case.

15.4   IRREDUCIBLE MATRIX REPRESENTATIONS

This section introduces irreducible representations, an enumerable set of important representations which exist for any finite group. In analogy with the basis vectors of a linear space, these irreducible representations are orthogonal and complete. This means, among other things, that any representation can be written as a sum of the irreducible representations, a fact that has great consequences in the application of group theory to physical problems.

The derivation of many of the results of this section, including the comments in the previous paragraph, are too involved to present in this short introduction to group theory. In addition, there are many facets of the subject that we will not discuss. Only a few important results are presented here. The complete story, along with formal proofs, can be found in many textbooks devoted solely to group theory. An introductory, but complete treatise of group representation theory can be found in the third chapter of the book by Wu-Ki Tung. The books by Hamermesh and Wigner also give a very complete discussion of group theory.

15.4.1   Notation

Because there are many possible matrix representations for a group, we need a notation for distinguishing them. When we discussed the group C_4, we explored the two different representations given in Table 15.3 and Table 15.6. We will call the first $C_4^{[2]}$ and the second $C_4^{[A]}$, as shown in Table 15.17. The brackets are used as a reminder that this notation refers to *matrix* representations only. You will discover the justification for labeling some representations with bracketed numbers and some with bracketed letters in the sections that follow. Also shown in Table 15.17 is the representation $C_4^{[1]}$, a one-dimensional representation made up of all 1's. This trivial representation will work for any group, and we will always reserve the bracketed index [1] for it.

TABLE 15.17. Three C_4 Representations

	I	g_2	g_3	g_4
$C_4^{[1]}$	1	1	1	1
$C_4^{[2]}$	$e^{i0} = 1$	$e^{i\pi/2} = i$	$e^{i\pi} = -1$	$e^{i3\pi/2} = -i$
$C_4^{[A]}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

TABLE 15.18. The Direct Sum of Two C_4 Representations

	I	g_2	g_3	g_4
$C_4^{[2]}$	1	i	-1	$-i$
$C_4^{[A]}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$
$C_4^{[B]} = C_4^{[2]} \oplus C_4^{[A]}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} i & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$	$\begin{bmatrix} -i & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$

To refer to a single element in a particular matrix representation, we use a modification of this notation. For example, the $C_4^{[A]}$ matrix representation of the g_3 group element is written as $g_3^{[A]}$. If we need to refer to one arbitrary element, we use a letter as the subscript. For example, $g_i^{[A]}$ refers to any one of the matrices in the $C_4^{[A]}$ representation.

15.4.2   Direct Sums of Matrix Representations

The matrix representation of a group can be of any dimension. In fact, given any matrix representation of a group, it is easy to expand it to a higher dimension by adding "blocks" of another representation. This process can be demonstrated using the C_4 cyclic group. Consider the $C_4^{[2]}$ and the $C_4^{[A]}$ representations shown in the first two rows of Table 15.18. The $C_4^{[2]}$ representation can be looked at as a set of 1×1 matrices. They can be added to the $C_4^{[A]}$ matrices in a block diagonal sense to form the new 3×3 matrix representation listed in the last row of Table 15.18. The $\oplus$ symbol, called the *direct sum* operation, is used to indicate this type of block diagonal addition:

$$C_4^{[B]} = C_4^{[2]} \oplus C_4^{[A]} \equiv \begin{bmatrix} C_4^{[2]} & 0 \\ 0 & C_4^{[A]} \end{bmatrix}. \tag{15.31}$$

It is easy to see that the set of $C_4^{[2]} \oplus C_4^{[A]}$ matrices form a valid representation for the C_4 group. When any two of these elements are multiplied together, the block diagonal form insures there is no mixing between blocks. Since each block independently satisfies the C_4 group multiplication table, the direct sum must also satisfy the same table.

15.4.3   Reducible and Irreducible Representations

Given a particular matrix representation $G^{[A]}$ for a group G,

$$G^{[A]} = \left\{ g_1^{[A]}, \; g_2^{[A]}, \; g_3^{[A]}, \; \cdots \; g_h^{[A]} \right\}, \tag{15.32}$$

we can generate another representation $G^{[B]}$ by simply transforming the coordinate basis. If $[T]$ is the transformation matrix, then the second representation is related to the first via the relations

$$g_i^{[B]} = [T]^{-1} g_i^{[A]} [T].$$

(15.33)

Keep in mind, the same transformation matrix $[T]$ must be applied to all the $g_i^{[A]}$ elements. It is obvious that both representations are of the same dimension and obey the same multiplication table. Notice Equation 15.33 is in the form of a similarity transformation, like that of Equation 15.23, so the matrices of the new representation will have the same character as the original representation.

Two matrix representations are said to be *equivalent* if you can obtain one from the other using the type of operation described in Equation 15.33. In contrast, *inequivalent* representations cannot be generated from one another by simply transforming the coordinate basis. Obviously, two representations of different dimension will always be inequivalent, but it is also possible to have two inequivalent representations of the same dimension.

If, using the same transformation matrix, all elements of a representation can be put into the form of a direct sum of two or more smaller blocks, the original representation is said to be *reducible*. If it cannot, it is *irreducible*. The irreducible representations of a group are important because any representation can be transformed to a direct sum of one or more irreducible representations. We will discuss this in more detail, later in this chapter.

It turns out that the number of inequivalent irreducible representations is not infinite, but rather equal to the number of different classes in the group. In order to distinguish these enumerable irreducible representations from the infinite possible reducible ones, we use a slightly different notation for the two. If a representation is reducible, we use a letter in the brackets. For example, the reducible representation of C_4 we developed in Equation 15.31 was called $C_4^{[B]}$. The representation $C_4^{[2]}$ is obviously irreducible (because it is one-dimensional) and so is labeled with a number.

Example 15.3 As a simple example of the reduction of a representation, consider the $C_4^{[A]}$ representation of the C_4 cyclic group, which was shown in Table 15.17. We would like to know if, using some similarity transformation, can we write this representation as the direct sum of two 1×1 matrices. In other words, is this representation reducible?

Since these matrices are two-dimensional, to reduce them any further is equivalent to diagonalizing them. In Chapter 4, we discussed a general method for finding the transformation matrix which diagonalizes a given matrix. Let's apply this method to find the matrix $[T]$ which diagonalizes the $g_2^{[A]}$ matrix element of $C_4^{[A]}$. First, we find the eigenvalues using the determinant equation

$$\begin{vmatrix} -\lambda & 1 \\ -1 & -\lambda \end{vmatrix} = 0,$$

(15.34)

TABLE 15.19. Block Diagonalization of $C_4^{[A]}$

I	g_2	g_3	g_4
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$	$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$	$\begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}$

which gives

$$\lambda = \pm i. \tag{15.35}$$

These eigenvalues generate the two eigenvectors

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix} \quad ; \quad \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix}. \tag{15.36}$$

From these we can construct the transformation matrix

$$[T] = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \tag{15.37}$$

and its inverse

$$[T]^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}. \tag{15.38}$$

Because we constructed it specifically to do so, applying $[T]$ and $[T]^{-1}$ to the $g_2^{[A]}$ matrix in a similarity transformation will certainly diagonalize it. The real test, however, is if the same transformation simultaneously diagonalizes all the other elements. Table 15.19 shows the result when the same similarity transformation is applied to the rest of the matrix elements. Notice, these elements are also diagonalized, so $C_4^{[A]}$ is reducible and can be written as the direct sum

$$C_4^{[2]} \oplus C_4^{[3]}, \tag{15.39}$$

where $C_4^{[3]}$ is the new representation shown in Table 15.20. Since $C_4^{[3]}$ is one-dimensional, and is obviously different from $C_4^{[1]}$ and $C_4^{[2]}$, it must be a third irreducible representation of C_4.

Earlier we stated that the number of irreducible representations is equal to the number of classes. The number of classes of C_4 is four, but so far we have only identified three different irreducible representations. One way to discover the fourth

TABLE 15.20. Another Irreducible Representation for C_4

	I	g_2	g_3	g_4
$C_4^{[3]}$	1	$-i$	-1	i

set would be by trial and error. A more educational method is to derive it using the orthogonality relations of the irreducible representations, a topic we explore later in this chapter.

Example 15.4   As another example of representation reduction, consider the D_3 group. We showed a 2×2 matrix representation for this group in Table 15.10. An attempt to diagonalize this representation in the same way we did for the previous example fails. Therefore, it is an irreducible representation, which we will call $D_3^{[3]}$. As always, $D_3^{[1]}$ is the trivial one-dimensional irreducible representation.

A 3×3 matrix representation for this group can be constructed from the operator representation developed in Table 15.12. For example, take the [231] operator. It generates the sequence (bca) from the sequence (abc). This operation can also be accomplished by a 3×3 matrix

$$
\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} b \\ c \\ a \end{bmatrix}. \tag{15.40}
$$

This matrix is a valid representation for the g_3 element of D_3. The matrices for all the other group elements are constructed in a similar manner. The result is shown in Table 15.21 and we will refer to this as the $D_3^{[A]}$ representation. Now we ask, is this representation reducible?

We can attempt to reduce this matrix by a similar process to the one used earlier. We will diagonalize one of the group elements, and then check the other elements to see if the same transformation puts them into a reduced form. We have no guarantee that this will work in this case, because there is the possibility that none of the elements (besides the identity) are diagonal in the reduced representation. Later in the text, we will revisit this problem, using a more reliable technique based on the orthogonality and completeness properties of the irreducible representations.

We have a small hint from looking back at the 2×2 $D_3^{[3]}$ representation. Its g_6 element is diagonalized. Since the $D_3^{[3]}$ matrices are an irreducible representation, it would make sense to attempt diagonalization of the same element in the 3×3 $D_3^{[A]}$ representation, and see what happens. Luckily, this process actually works, and the transformed version of the set of $D_3^{[A]}$ matrices is given in Table 15.22. Notice this is just the direct sum $D_3^{[1]} \oplus D_3^{[3]}$.

TABLE 15.21. The $3 \times 3\, D_3^{[A]}$ Representation

I	g_2	g_3	g_4	g_5	g_6
$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

TABLE 15.22. Block Diagonalized of the $D_3^{[4]}$ Representation

I	g_2	g_3
$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \dfrac{-1}{2} & \dfrac{-\sqrt{3}}{2} \\ 0 & \dfrac{\sqrt{3}}{2} & \dfrac{-1}{2} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \dfrac{-1}{2} & \dfrac{\sqrt{3}}{2} \\ 0 & \dfrac{-\sqrt{3}}{2} & \dfrac{-1}{2} \end{bmatrix}$

g_4	g_5	g_6
$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \dfrac{1}{2} & \dfrac{-\sqrt{3}}{2} \\ 0 & \dfrac{-\sqrt{3}}{2} & \dfrac{-1}{2} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \dfrac{1}{2} & \dfrac{\sqrt{3}}{2} \\ 0 & \dfrac{\sqrt{3}}{2} & \dfrac{-1}{2} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

We already determined that D_3 has three different classes. This means that there should be three irreducible representations. So far, only two have been identified, $D_3^{[1]}$ and $D_3^{[3]}$. In the next section, we determine the remaining irreducible representation $D_3^{[2]}$, using the orthogonality of the irreducible representations.

15.4.4 Orthogonality of Irreducible Representations

There are several important rules that the irreducible matrix representations obey:

1. The number of irreducible matrix representations for a group is equal to the number of different classes of the group.

2. If $n^{[1]}$ is the dimension of the first irreducible matrix representation, $n^{[2]}$ the dimension of the second, and so on, then

$$(n^{[1]})^2 + (n^{[2]})^2 + (n^{[3]})^2 + \cdots = h, \tag{15.41}$$

where the summation is over all the irreducible representations and h is the order of the group.

3. "Vectors" formed by the elements of the irreducible representation matrices are orthogonal. In this context, a vector is formed by combining the elements of a representation in a very particular way. The components of one such vector, for example, are formed by all the 1-2 matrix elements of the six $D_3^{[3]}$ matrices.

The components of another are formed by all the 2-2 elements of the six $D_3^{[3]}$ matrices, etc.

4. The magnitude squared of each of the "vectors" described in rule #3 is equal to $h/n^{[i]}$ where $n^{[i]}$ is the dimension of the matrix representation used to construct the vector.

When checking orthogonality or checking the magnitude of the vectors described in rules #3 and #4, be aware that, if any of the components are complex, one set of the vector components should be complex conjugated before performing any dot products.

In both the examples of the previous section, we found there was one missing irreducible representation. Instead of using guesswork to find the last representation, it is instructive to use the above rules to determine it.

Example 15.5 For the C_4 cyclic group, three irreducible representations were identified: $C_4^{[1]}$, $C_4^{[2]}$, and $C_4^{[3]}$. There are four different classes in this group, and according to rule #1, there must be another irreducible representation which we will call $C_4^{[4]}$.

First let's determine the dimension of the $C_4^{[4]}$ matrices. According to rule #2:

$$(n^{[1]})^2 + (n^{[2]})^2 + (n^{[3]})^2 + (n^{[4]})^2 = 4 \tag{15.42}$$

$$1 + 1 + 1 + (n^{[4]})^2 = 4 \tag{15.43}$$

or

$$n^{[4]} = 1. \tag{15.44}$$

So the final irreducible representation for the C_4 group is composed of four 1×1 matrices. The identity element must be 1, and according to the multiplication table for C_4, g_3 is its own inverse, so $g_3^{[4]} = \pm 1$. Also, g_4 is the inverse of g_2, so if we let $g_2^{[4]} = a$, then $g_4^{[4]} = 1/a$. Notice we have given a the flexibility of being complex. This information leads us to the partially specified Table 15.23. Now we just need to determine the value of a and the sign of $g_3^{[4]}$.

A table of all of the four irreducible representations for C_4 is shown in Table 15.24. Notice all the representations are 1×1 matrices. Incidentally, this turns out to be a general property of any Abelian group. Following the prescription described in rule #3, we form four, four-dimensional vectors, which we call $\overline{V}^{[1]}$, $\overline{V}^{[2]}$, $\overline{V}^{[3]}$, and $\overline{V}^{[4]}$:

TABLE 15.23. The Form of the Missing Irreducible Representation of C_4

	I	g_2	g_3	g_4
$C_4^{[4]}$	1	a	± 1	$1/a$

TABLE 15.24. Partially Specified Table of the Irreducible Representations of C_4

	g_1	g_2	g_3	g_4
$C_4^{[1]}$	1	1	1	1
$C_4^{[2]}$	1	i	-1	$-i$
$C_4^{[3]}$	1	$-i$	-1	i
$C_4^{[4]}$	1	$\underline{a}$	± 1	$1/\underline{a}$

TABLE 15.25. The Irreducible Representations for C_4

	I	g_2	g_3	g_4
$C_4^{[1]}$	1	1	1	1
$C_4^{[2]}$	1	i	-1	-1
$C_4^{[3]}$	1	$-i$	-1	i
$C_4^{[4]}$	1	-1	1	-1

$$\overline{\mathbf{V}}^{[1]} \rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \overline{\mathbf{V}}^{[2]} \rightarrow \begin{bmatrix} 1 \\ i \\ -1 \\ -i \end{bmatrix} \quad \overline{\mathbf{V}}^{[3]} \rightarrow \begin{bmatrix} 1 \\ -i \\ -1 \\ i \end{bmatrix} \quad \overline{\mathbf{V}}^{[4]} \rightarrow \begin{bmatrix} 1 \\ \underline{a} \\ \pm 1 \\ 1/\underline{a} \end{bmatrix} \qquad (15.45)$$

According to the rule, these vectors must be orthogonal. If $\overline{\mathbf{V}}^{[4]}$ is orthogonal to $\overline{\mathbf{V}}^{[1]}$ with $g_3^{[4]} = -1$, we find that $\underline{a} = \pm i$. Either choice for the sign of $\underline{a}$ does not lead us to a representation of C_4 that we have not seen before. But if instead we pick $g_3^{[4]} = +1$, we obtain a new irreducible representation with $\underline{a} = -1$. Now our set of irreducible representations is complete. They are summarized in Table 15.25.

According to rule #4, the magnitudes of $\overline{\mathbf{V}}^{[1]}$, $\overline{\mathbf{V}}^{[2]}$, $\overline{\mathbf{V}}^{[3]}$, and $\overline{\mathbf{V}}^{[4]}$ should all be equal to 4. Taking the dot product of each vector with itself, remembering to take the complex conjugate of one of the terms, shows this is indeed true.

Example 15.6 The same procedure used above can be repeated for the D_3 group. This group is of order six, and has three classes. We have already identified two inequivalent irreducible representations: $D_3^{[1]}$ and $D_3^{[3]}$. That leaves a third, the $D_3^{[2]}$ representation, for us to find.

Using rule #2, the dimension of $D_3^{[2]}$ is given by

$$(n^{[1]})^2 + (n^{[2]})^2 + (n^{[3]})^2 = 6 \qquad (15.46)$$

$$1 + (n^{[2]})^2 + 4 = 6 \qquad (15.47)$$

TABLE 15.26. The Form of the Third Irreducible Representation for D_3

	g_1	g_2	g_3	g_4	g_5	g_6
$D_3^{[4]}$	1	1	1	± 1	± 1	± 1

or

$$(n^{[2]}) = 1. \tag{15.48}$$

So the irreducible representation we seek is a set of six 1×1 matrices. According to the multiplication table for D_3, the g_4, g_5 and g_6 elements are all their own inverses, and are also in the same class. This lets us write $g_4^{[2]} = g_5^{[2]} = g_6^{[2]} = \pm 1$. Also according to the multiplication table, $g_2 \, g_4 = g_5$ and $g_3 \, g_4 = g_6$, so we must have $g_2^{[2]} = g_3^{[2]} = +1$. These properties give us the partially specified representation in Table 15.26. If we take the positive sign for all these terms, we just get a repeat of the trivial representation $D_3^{[1]}$. Therefore we take the negative signs instead, and the complete set of irreducible representations for D_3 are listed in Table 15.27.

It is educational to check that the other group rules hold for these representations. Rule #3 specifies that we can form "vectors" from the elements of the matrices, and that these vectors should all be orthogonal to one another. There will be six of these vectors. The 1×1 $D_3^{[1]}$ representation forms the first vector:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}. \tag{15.49}$$

The second vector comes from the $D_3^{[2]}$ representation:

TABLE 15.27. The Irreducible Representations for D_3

	I	g_2	g_3	g_4	g_5	g_6
$D_3^{[1]}$	1	1	1	1	1	1
$D_3^{[2]}$	1	1	1	-1	-1	-1
$D_3^{[3]}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} \frac{-1}{2} & \frac{-\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{-1}{2} \end{bmatrix}$	$\begin{bmatrix} \frac{-1}{2} & \frac{\sqrt{3}}{2} \\ \frac{-\sqrt{3}}{2} & \frac{-1}{2} \end{bmatrix}$	$\begin{bmatrix} \frac{1}{2} & \frac{-\sqrt{3}}{2} \\ \frac{-\sqrt{3}}{2} & \frac{-1}{2} \end{bmatrix}$	$\begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{-1}{2} \end{bmatrix}$	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}. \tag{15.50}$$

The 2×2 $D_3^{[3]}$ representation generates the last four:

$$\begin{bmatrix} 1 \\ -1/2 \\ -1/2 \\ 1/2 \\ 1/2 \\ -1 \end{bmatrix} \begin{bmatrix} 0 \\ -\sqrt{3}/2 \\ \sqrt{3}/2 \\ -\sqrt{3}/2 \\ \sqrt{3}/2 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ \sqrt{3}/2 \\ -\sqrt{3}/2 \\ -\sqrt{3}/2 \\ \sqrt{3}/2 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1/2 \\ -1/2 \\ -1/2 \\ -1/2 \\ 1 \end{bmatrix}. \tag{15.51}$$

Taking the dot product of all fifteen possible pairs of these vectors shows they are indeed orthogonal. According to rule #4, the vectors which come from $D_3^{[1]}$ and $D_3^{[2]}$ should have magnitude squared values of 6 and those from $D_3^{[3]}$ should have 3. This is also obviously true.

15.4.5 Character Tables and Character Orthogonality

Once all the irreducible representations are known for a particular group, it is useful to construct a *character table*. This table lists the character of each group element for each of the irreducible representations. A character table for the D_3 group is shown in Table 15.28.

Similar to the orthogonality rules presented in the previous section, the character tables have orthogonality rules of their own. In this case, the "vectors" formed by the columns of these tables are orthogonal to one another, while their magnitudes squared are equal to the order of the group. It is easy to check that these rules hold for the D_3 character table.

TABLE 15.28. A Character Table for the Irreducible Representations of D_3

D_3	$\chi^{[1]}$	$\chi^{[2]}$	$\chi^{[3]}$
g_1	1	1	2
g_2	1	1	-1
g_3	1	1	-1
g_4	1	-1	0
g_5	1	-1	0
g_6	1	-1	0

TABLE 15.29. A More Compact Character Table for D_3

D_3	$\chi^{[1]}$	$\chi^{[2]}$	$\chi^{[3]}$
I	1	1	2
g_2, g_3	1	1	-1
g_4, g_5, g_6	1	-1	0

Since elements in the same class always have the same character, the character table is often written more concisely by putting all the elements in the same class in a single row. The compact character table for D_3 is given in Table 15.29. When forming dot products between the columns of these character tables, you need to weight each term by the number of elements in each class.

15.4.6   Completeness of the Irreducible Representations

The real power of the irreducible representations stems from the fact that, using the proper similarity transformation, *any* matrix representation can be written as a direct sum of one or more of the irreducible representations. A particular irreducible representation may appear once, more than once, or not at all. In other words, with some $[T]$, any group representations $G^{[a]}$, can be written

$$[T]^{-1} G^{[a]} [T] = G^{[1]} \oplus G^{[2]} \oplus \cdots . \tag{15.52}$$

This is often called the *decomposition* of a representation.

There is a remarkable result of group theory which says that the number of times a particular irreducible representation appears in the direct sum of Equation 15.52 can be determined just by looking at the character table. Take the vector formed by the characters of the representation you want to decompose. Then form the dot product of this vector with one formed from the character values of a particular irreducible representation. Divide the value of this dot product by the order of the group, and the result will be equal to the number of times the particular irreducible representation appears in the decomposition. Notice how the irreducible representations are playing a role analogous to that of the basis vectors of a coordinate system. Because all matrix representations can be expanded this way, the irreducible representations are said to be *complete*.

Example 15.7   In a previous example, we determined that the 3×3 $D_3^{[A]}$ matrix representation in Table 15.21 could be transformed and written as the direct sum $D_3^{[1]} \oplus D_3^{[3]}$. We arrived at the result using the diagonalization process and a little guesswork. Let's see how the decomposition rule described above gives us this result with much less work. The character vector for the Table 15.21 representation for $D_3^{[A]}$

is

$$\overline{\mathbf{V}} \rightarrow \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}. \tag{15.53}$$

Now take the dot product of this vector with each of the three vectors of the D_3 character table listed in Table 15.28, and divide by 6. The net result is that there is one $D_3^{[1]}$ representation, one $D_3^{[2]}$ representation, and no $D_3^{[3]}$ representation in the block diagonalization. This is exactly the result we found before.

15.4.7 Shur's Lemma

Many practical applications of matrix representations center around an important group theory result known as *Shur's lemma*. Imagine we have a group G, and $G^{[x]}$ is one of its matrix representations:

$$G^{[x]} = \left\{ g_1^{[x]}, g_2^{[x]}, \cdots g_h^{[x]} \right\}. \tag{15.54}$$

Now suppose we have a matrix $[V]$ which is invariant under similarity transformations by *all* these matrix elements of $G^{[x]}$. That is,

$$g_i^{[x]^{-1}} [V] g_i^{[x]} = [V] \quad \text{for } i = 1, 2, \cdots h. \tag{15.55}$$

Shur's lemma says if $[V]$ satisfies Equation 15.55, and $G^{[x]}$ is an irreducible representation, then $[V]$ must be a multiple of the identity matrix. That is,

$$[V] = \lambda [I], \tag{15.56}$$

where λ is some constant. If $[V]$ satisfies the conditions of Equation 15.55, it is said to *commute* with the elements of $G^{[x]}$. This terminology is used because an equivalent way of writing Equation 15.55 is

$$[V] g_i^{[x]} = g_i^{[x]} [V]. \tag{15.57}$$

If $G^{[x]}$ is reducible, we cannot use Shur's lemma directly. But remember from our previous discussion, any reducible representation can be decomposed into the direct sum of two or more irreducible representations by using the appropriate coordinate transformation. For example, imagine the decomposition of $G^{[x]}$ looks like

$$[T]^{-1} G^{[x]} [T] = \begin{bmatrix} G^{[1]} & 0 & 0 \\ 0 & G^{[2]} & 0 \\ 0 & 0 & G^{[3]} \end{bmatrix}, \tag{15.58}$$

where $G^{[1]}$, $G^{[2]}$, and $G^{[3]}$ are irreducible representations of dimension 1, 2, and 4 respectively, and $[T]$ is the transformation matrix that puts the matrices into the block diagonal form. If $[V]$ commutes with all the elements of $G^{[x]}$, the transformed matrix $[V'] \equiv [T]^{-1}[V][T]$ will commute with all the elements of the transformed $G^{[x]}$ representation. Then, by Shur's lemma, the $[V]'$ matrix must be diagonal, with the form

$$[V'] = [T]^{-1}[V][T] = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 \end{bmatrix} \qquad (15.59)$$

There are three different constants along the diagonal of $[V']$, a different one for each block of the decomposition of $G^{[x]}$. This works because Shur's lemma applies separately to each irreducible block.

Example 15.8 We now present an example that shows how many of the ideas of group theory can be applied to a problem from classical mechanics. Consider the two-dimensional system of balls and springs shown in Figure 15.9. Each of the three balls has mass m, and the three springs have the same spring constant k. What are the normal frequencies of vibration for this system? Because of the symmetry, we should expect group theory to be useful in solving this problem. We begin with a review of the standard method of determining the normal modes of such a system and then show how group theory simplifies the process.

Since there are three masses and we are working in two dimensions, we need a total of six coordinates to describe an arbitrary state of the system. We will use the coordinates q_1 and q_2 to describe the displacement of the mass in the lower-left
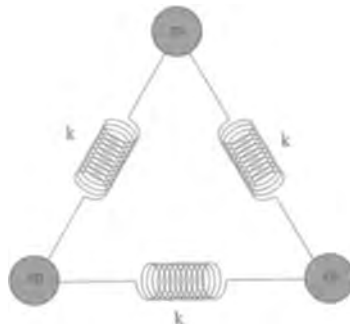
Figure 15.9 Coupled Mass Spring System
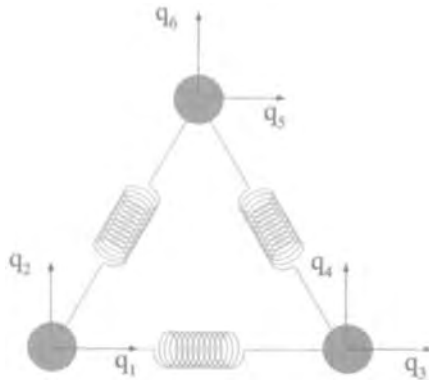
Figure 15.10 Coordinates for the Mass Spring Problem

corner relative to its equilibrium position, as shown in Figure 15.10. Likewise, there is one pair of coordinates for each of the other two masses shown in the figure. If we assume the distance between the masses is much greater than the magnitude of the displacements, the potential energy stored in the springs can be written as

$$V = \frac{1}{2}k \begin{bmatrix} (q_3 - q_1)^2 \\ + \left[\frac{-1}{2}(q_5 - q_3) + \frac{\sqrt{3}}{2}(q_6 - q_4) \right]^2 \\ + \left[\frac{1}{2}(q_1 - q_5) + \frac{\sqrt{3}}{2}(q_2 - q_6) \right]^2 \end{bmatrix}. \tag{15.60}$$

It is convenient to arrange the potential energy of Equation 15.60 into the form of a matrix $[V]$, which is defined so that

$$V = \frac{1}{2}kV_{ij}q_iq_j, \tag{15.61}$$

where there is an implicit sum over both i and j. A bit of algebraic manipulation of Equation 15.60 gives the $[V]$ matrix as

$$[V] = \begin{bmatrix} 5/4 & \sqrt{3}/4 & 1 & 0 & -1/4 & -\sqrt{3}/4 \\ \sqrt{3}/4 & 3/4 & 0 & 0 & -\sqrt{3}/4 & -3/4 \\ 1 & 0 & 5/4 & -\sqrt{3}/4 & -1/4 & \sqrt{3}/4 \\ 0 & 0 & -\sqrt{3}/4 & 3/4 & \sqrt{3}/4 & -3/4 \\ -1/4 & -\sqrt{3}/4 & -1/4 & \sqrt{3}/4 & 1/2 & 0 \\ -\sqrt{3}/4 & -3/4 & \sqrt{3}/4 & -3/4 & 0 & 3/2 \end{bmatrix}. \tag{15.62}$$

To determine the equations of motion, we can use Newton's second law to write

$$m\frac{d^2q_i}{dt^2} = -\frac{\partial V}{\partial q_i} = -kV_{ij}q_j. \tag{15.63}$$

A normal mode is a state of motion that oscillates with constant frequency. To find the normal modes, we therefore require all the coordinates oscillate at the same frequency, i.e., $q_i \rightarrow \underline{q}_i e^{i\omega t}$. Applying this to Equation 15.63 gives

$$\frac{m\omega^2}{k}\underline{q}_i = V_{ij}\underline{q}_j. \qquad (15.64)$$

But notice this is the same as a typical eigenvalue problem,

$$V_{ij}\underline{q}_j = \lambda\underline{q}_i, \qquad (15.65)$$

where the eigenvalues are related to the normal frequencies by

$$\lambda = \frac{m\omega^2}{k}. \qquad (15.66)$$

This means that to find the normal frequencies we just need to diagonalize the $[V]$ matrix. Unfortunately, diagonalizing a typical six-dimensional matrix is no easy task, since it involves finding the roots of a sixth-order polynomial equation. But the underlying symmetry of this problem, plus some physical reasoning, will let us find these eigenvalues much more easily.

The physical system clearly has a D_3-type symmetry. The representation of D_3 we need to consider for this problem is the set of 6×6 transformation matrices that correspond to each symmetry operation acting on the q_i coordinates. Let's work out the g_2 element of this matrix representation in detail. Looking back at Table 15.10, this element corresponds to a rotation of $2\pi/3$ counterclockwise around a z-axis sticking out of the page of Figure 15.10. Consequently, this operation rotates the q_i coordinates into the q_i' coordinates as shown in Figure 15.11. We can write the
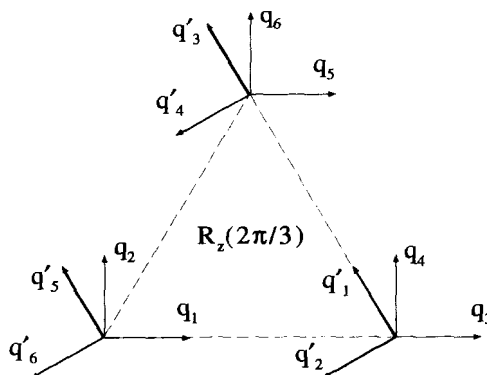
Figure 15.11   Coordinate Transformation for $R_z(2\pi/3)$ Element

relationships between the primed and unprimed coordinates as

$$q_1 = -\frac{1}{2}q_5' - \frac{\sqrt{3}}{2}q_6' \quad q_2 = \frac{\sqrt{3}}{2}q_5' - \frac{1}{2}q_6'$$

$$q_3 = -\frac{1}{2}q_1' - \frac{\sqrt{3}}{2}q_2' \quad q_4 = \frac{\sqrt{3}}{2}q_1' - \frac{1}{2}q_2'$$

$$q_5 = -\frac{1}{2}q_3' - \frac{\sqrt{3}}{2}q_4' \quad q_6 = \frac{\sqrt{3}}{2}q_3' - \frac{1}{2}q_4', \quad (15.67)$$

so the matrix which transforms the primed coordinates into the unprimed coordinates is

$$R_z(2\pi/3) \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & -1/2 & -\sqrt{3}/2 \\ 0 & 0 & 0 & 0 & \sqrt{3}/2 & -1/2 \\ -1/2 & -\sqrt{3}/2 & 0 & 0 & 0 & 0 \\ \sqrt{3}/2 & -1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1/2 & -\sqrt{3}/2 & 0 & 0 \\ 0 & 0 & \sqrt{3}/2 & -1/2 & 0 & 0 \end{bmatrix}.$$

$$(15.68)$$

Repeating this process for the other D_3 elements generates the complete representation, given in Table 15.30. In addition, this table lists the character of each matrix.

These matrices are reducible, and we can use the completeness and orthogonality of the character tables to determine how many times each irreducible representation of D_3 occurs. Looking at Table 15.30, the character vector for this representation is

$$\begin{bmatrix} 6 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (15.69)$$

If we form the dot product with the columns of the irreducible character table for D_3 given in Table 15.28, we find the decomposition for the 6×6 representation is:

$$D_3^{[1]} \oplus D_3^{[2]} \oplus D_3^{[3]} \oplus D_3^{[3]}. \quad (15.70)$$

Because the potential matrix must obey the underlying symmetry of the system, it must be invariant to similarity transformations performed by all of the elements of the 6×6 representation for D_3. By Shur's lemma, the transformation matrix $[T]$ that

TABLE 15.30. 6 × 6 Matrix Representation and Character Values for the Spring Problem

	g	χ
I	$R(0) \rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	6
g_2	$R_z(2\pi/3) \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & -1/2 & -\sqrt{3}/2 \\ 0 & 0 & 0 & 0 & \sqrt{3}/2 & -1/2 \\ -1/2 & -\sqrt{3}/2 & 0 & 0 & 0 & 0 \\ \sqrt{3}/2 & -1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1/2 & -\sqrt{3}/2 & 0 & 0 \\ 0 & 0 & \sqrt{3}/2 & -1/2 & 0 & 0 \end{bmatrix}$	0
g_3	$R_z(4\pi/3) \rightarrow \begin{bmatrix} 0 & 0 & -1/2 & \sqrt{3}/2 & 0 & 0 \\ 0 & 0 & -\sqrt{3}/2 & -1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1/2 & \sqrt{3}/2 \\ 0 & 0 & 0 & 0 & -\sqrt{3}/2 & -1/2 \\ -1/2 & \sqrt{3}/2 & 0 & 0 & 0 & 0 \\ -\sqrt{3}/2 & -1/2 & 0 & 0 & 0 & 0 \end{bmatrix}$	0
g_4	$R_D(\pi) \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & 1/2 & -\sqrt{3}/2 \\ 0 & 0 & 0 & 0 & -\sqrt{3}/2 & -1/2 \\ 0 & 0 & 1/2 & -\sqrt{3}/2 & 0 & 0 \\ 0 & 0 & -\sqrt{3}/2 & -1/2 & 0 & 0 \\ 1/2 & -\sqrt{3}/2 & 0 & 0 & 0 & 0 \\ -\sqrt{3}/2 & -1/2 & 0 & 0 & 0 & 0 \end{bmatrix}$	0
g_5	$R_E(\pi) \rightarrow \begin{bmatrix} 1/2 & \sqrt{3}/2 & 0 & 0 & 0 & 0 \\ \sqrt{3}/2 & -1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & \sqrt{3}/2 \\ 0 & 0 & 0 & 0 & \sqrt{3}/2 & -1/2 \\ 0 & 0 & 1/2 & \sqrt{3}/2 & 0 & 0 \\ 0 & 0 & \sqrt{3}/2 & -1/2 & 0 & 0 \end{bmatrix}$	0
g_6	$R_F(\pi) \rightarrow \begin{bmatrix} 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	0

block diagonalizes the 6×6 representation will place the $[V]$ matrix in the form

$$[V'] = [T]^{-1}[V][T] = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_4 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_4 \end{bmatrix}, \tag{15.71}$$

where we have used $[V']$ to represent the $[V]$ matrix in this diagonalizing coordinate system. Already group theory has reduced the number of unknown eigenvalues of the $[V]$ matrix, and therefore the number of normal mode frequencies for our mass spring problem, from six to four. Notice that this has been accomplished without actually finding the elements of the $[T]$ matrix.

The next simplification comes not from group theory, but rather from some general physical reasoning. When finding the normal modes of any unconstrained system, there will always be a number of modes with zero frequency. These correspond to rigid displacements of the system. In a two-dimensional system, such as this example, there must be exactly three such modes: two translational and one rotational. We can pick λ_1 and the two λ_3's to be the eigenvalues associated with these modes, which simplifies the $[V']$ matrix to

$$[V'] = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_4 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_4 \end{bmatrix}. \tag{15.72}$$

Now the problem has only two unknowns.

We can determine the values of the remaining eigenvalues by remembering that the trace of a matrix is invariant to similarity transformations. Therefore, the trace of the original $[V]$ matrix must equal the trace of $[V']$:

$$Tr[V] = Tr[V']$$
$$6 = \lambda_2 + 2\lambda_4. \tag{15.73}$$

There is only one free parameter left, and we need one more condition to specify things completely. The simplest way to obtain it is to consider the squares of the potential matrices: $[V]^2$ and $[V']^2$. The trace of these two matrices must also be unaffected by similarity transformations, so we can write

$$Tr\left([V]^2\right) = Tr\left([V']^2\right)$$
$$\frac{27}{2} = \lambda_2{}^2 + 2\lambda_4{}^2. \tag{15.74}$$

The simultaneous solution of Equations 15.73 and 15.74 gives the results $\lambda_2 = 3$ and $\lambda_4 = 3/2$.

Combining these results with Equation 15.66, we finally have the normal frequencies:

$$\omega_a = 0 \qquad \omega_b = \sqrt{\frac{3}{2}}\sqrt{\frac{k}{m}} \qquad \omega_c = \sqrt{3}\sqrt{\frac{k}{m}}. \qquad (15.75)$$

The first two frequencies are degenerate, since ω_a occurs three times and ω_b twice.

15.5 CONTINUOUS GROUPS

So far, all the groups we have discussed have had a finite number of discrete elements. There is another type of a group that has an infinite number of elements, described as a function of one or more continuous variables. This section briefly discusses three examples of this type of group.

15.5.1 The 2D Rotation Group

The C_4 group consisted of four elements that could be viewed physically as the rotations in a plane which left a square in a state indistinguishable from its initial state. We mentioned that this group could be generalized to C_n, to handle the symmetry of a polygon with n sides.

What happens when we let $n \to \infty$? Now we really have an infinite number of elements which correspond to any rotation angle in the plane. If we use the "R" notation for rotations, the group elements can be written as continuous function of the angle, which we call θ,

$$\lim_{n \to \infty} C_n \to R(\theta), \qquad (15.76)$$

with $0 \le \theta < 2\pi$. Let's check to be sure that this definition of the group elements satisfies the requirements of a group:

1. The two rotations applied in succession, $R(\theta_1)R(\theta_2)$, is equivalent to $R(\theta_1 + \theta_2)$, which is also an element of the group. Keep in mind that if $\theta_1 + \theta_2$ falls outside the range of 0 to 2π, it is necessary to subtract an integer multiple of 2π to put the result back into the 0 to 2π range.

2. The multiplication is associative, because $[R(\theta_1)R(\theta_2)]R(\theta_3) = R(\theta_1 + \theta_2 + \theta_3) = R(\theta_1)[R(\theta_2)R(\theta_3)]$.

3. There is an identity element, $I = R(0)$.

4. Each element $R(\theta)$ has an inverse $R(2\pi - \theta)$.

We can also generate some matrix representations for these elements. An obvious one-dimensional choice is the complex exponentials:

$$R(\theta) \rightarrow e^{i\theta}. \tag{15.77}$$

A different, two-dimensional representation is the set of matrices

$$R(\theta) \rightarrow \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}. \tag{15.78}$$

It is a worthwhile exercise to convince yourself that both these representations obey the underlying group multiplication table.

15.5.2 The 3D Rotation Group O_3^+

The set of all possible rotations in three dimensions is another example of a continuous group. It is easy to show that all four group axioms are still satisfied when we move into three dimensions.

The matrix form for the continuous two-dimensional rotation group, given by Equation 15.78, can be generalized to three dimensions. If we restrict the rotation to be around the z-axis, as shown in Figure 15.12, the matrix representation of the rotation is given by the 3×3 matrix

$$R_z(\theta) \rightarrow \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad 0 \le \theta < 2\pi. \tag{15.79}$$

An arbitrary orientation in three dimensions can be accomplished using successive rotations around *any* three different axes. There are many ways to accomplish this, but the conventional choice is the rotation sequence

$$R(\psi, \theta, \phi) = R_{z''}(\psi)R_{x'}(\theta)R_z(\phi). \tag{15.80}$$

Here the x'-axis is what the x-axis has become after the first rotation, and the z''-axis is what the z-axis has become after the second rotation. It is useful to also describe this process in words. First, rotate around the z-axis by an angle ϕ. This changes the
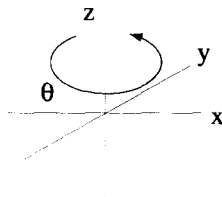
Figure 15.12 Rotation Only About the z-Axis
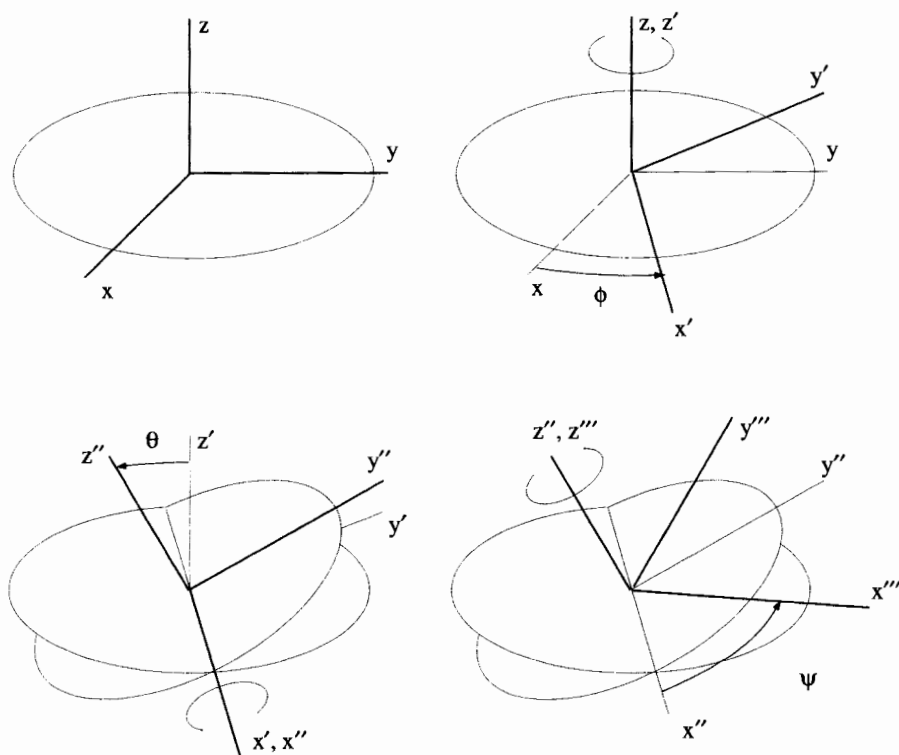
Figure 15.13   Euler's Angles

x-axis into a new axis, x', but leaves the z-axis alone. Next, rotate around this x'-axis by an amount θ. This transforms the z'-axis into the z''-axis. Finally, rotate around the z''-axis with angle ψ. The three angles ϕ, θ, and ψ are called the *Euler angles* and are shown pictorially in Figure 15.13.

We can write down the representations of each of the individual rotations that make up Equation 15.80. The rotation about z is the matrix

$$R_z(\phi) \rightarrow \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{15.81}$$

The rotation around x' in terms of the primed coordinates is given by

$$R_{x'}(\theta) \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}. \tag{15.82}$$

The final rotation around z'' in terms of the double primed coordinates is

$$R_{z''}(\psi) \rightarrow \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad (15.83)$$

Multiplication of these three matrices gives the intimidating 3×3 matrix:

$$R(\psi, \theta, \phi) = \begin{bmatrix} \begin{array}{c} -\cos\theta\sin\psi\sin\phi \\ +\cos\phi\cos\psi \end{array} & \vdots & \begin{array}{c} \cos\theta\sin\psi\cos\phi \\ +\sin\phi\cos\psi \end{array} & \vdots & \sin\theta\sin\psi \\ \\ \begin{array}{c} -\cos\theta\cos\psi\sin\phi \\ -\sin\psi\cos\phi \end{array} & \vdots & \begin{array}{c} \cos\theta\cos\psi\cos\phi \\ -\sin\psi\sin\phi \end{array} & \vdots & \sin\theta\cos\psi \\ \\ \sin\theta\sin\phi & \vdots & -\sin\theta\cos\phi & \vdots & \cos\theta \\ & \vdots & & \vdots & \end{bmatrix}.$$

$$(15.84)$$

Because this matrix represents a rigid rotation of coordinates, it is necessarily orthonormal. It also has a determinant of $+1$. In fact, if ψ, θ, and ϕ are allowed to roam over their entire 0 to 2π ranges, we have *all* the possible 3×3 orthogonal matrices with $+1$ determinants. For this reason, when the group of three-dimensional rotations is written in this manner, it is called O_3^+, where the O stands for orthogonal, the 3 for the dimension, and the $+$ to indicate the sign of the determinant. In the literature, it is also frequently called $SO(3)$, where the S stands for the "special" condition of a $+1$ determinant. You also might run into the term *unimodular* to describe matrices with $+1$ determinants.

15.5.3 The Group of Special Unitary Matrices $SU(2)$

Our last example of a continuous group is the complete set of 2×2 unitary matrices, with determinants of $+1$. This group is commonly called $SU(2)$. A matrix is unitary if its inverse is equal to the complex conjugate of its transpose. The general form of a 2×2 matrix that satisfies these conditions is

$$\begin{bmatrix} \underline{a} & \underline{b} \\ -\underline{b}^* & \underline{a}^* \end{bmatrix}, \qquad (15.85)$$

with the requirement that $\underline{a}\,\underline{a}^* + \underline{b}\,\underline{b}^* = 1$ to ensure the unimodular condition. Notice there are really three free parameters here, because the real and complex parts of $\underline{a}$ and $\underline{b}$ provide two each, but the unimodular condition reduces this number by one.

To show that these matrices form a group, we need to prove that the multiplication of any two of them still has the general form of Equation 15.85. This can be verified

by forming the product

$$\begin{bmatrix} \underline{a}_1 & \underline{b}_1 \\ -\underline{b}_1{}^* & \underline{a}_1^* \end{bmatrix} \begin{bmatrix} \underline{a}_2 & \underline{b}_2 \\ -\underline{b}_2^* & \underline{a}_2^* \end{bmatrix} = \begin{bmatrix} \underline{a}_1\underline{a}_2 - \underline{b}_1\underline{b}_2^* & \underline{a}_1\underline{b}_2 + \underline{b}_1\underline{a}_2^* \\ -\underline{a}_2\underline{b}_1^* - \underline{a}_1^*\underline{b}_2^* & -\underline{b}_1^*\underline{b}_2 + \underline{a}_1^*\underline{a}_2^* \end{bmatrix}. \tag{15.86}$$

Now define $\underline{a}_3 \equiv \underline{a}_1\underline{a}_2 - \underline{b}_1\underline{b}_2^*$ and $\underline{b}_3 \equiv \underline{a}_1\underline{b}_2 + \underline{b}_1\underline{a}_2^*$, and notice the right-hand side of Equation 15.86 can now be written as

$$\begin{bmatrix} \underline{a}_3 & \underline{b}_3 \\ -\underline{b}_3^* & \underline{a}_3^* \end{bmatrix}. \tag{15.87}$$

The unimodular condition still holds:

$$\begin{aligned}
\underline{a}_3\underline{a}_3^* + \underline{b}_3\underline{b}_3^* &= (\underline{a}_1\underline{a}_2 - \underline{b}_1\underline{b}_2^*)(\underline{a}_1^*\underline{a}_2 - \underline{b}_1^*\underline{b}_2^*) + (\underline{a}_1\underline{b}_2 + \underline{b}_1\underline{a}_2^*)(\underline{a}_1^*\underline{b}_2^* + \underline{b}_1^*\underline{a}_2) \\
&= \underline{a}_1\underline{a}_1^*\underline{a}_2\underline{a}_2^* + \underline{a}_1\underline{a}_1^*\underline{b}_2\underline{b}_2^* + \underline{b}_1\underline{b}_1^*\underline{b}_2\underline{b}_2^* + \underline{b}_1\underline{b}_1^*\underline{a}_2\underline{a}_2^* \\
&= (\underline{a}_1\underline{a}_1^* + \underline{b}_1\underline{b}_1^*)(\underline{a}_2\underline{a}_2^* + \underline{b}_2\underline{b}_2^*) \\
&= 1. \tag{15.88}
\end{aligned}$$

This group can be generalized to $SU(n)$, the set of $n \times n$ unimodular, unitary matrices. These matrices can be shown to have $(n^2 - 1)$ independent parameters.

EXERCISES FOR CHAPTER 15

1. Let S be a subgroup of a larger group G. If g_i is an element in G that is not in S, show that g_i^{-1} is also not in S.

2. Show that there are only two distinct fourth-order groups.

3. How many classes does the vierergruppe have?

4. Identify all the subgroups of D_3. For each subgroup, indicate the elements of all its possible left and right cosets. Show that the ratio h/h' is an integer for all of the subgroups, where h is the order of D_3 and h' is the order of the subgroup.

5. Three elements of a group are represented by the following 2×2 matrices,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} \underline{w} & 0 \\ 0 & \underline{w}^2 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

where $\underline{w} = e^{i2\pi/3}$.

(a) What are the other elements of this group?

(b) For each element identify its inverse.

(c) Find all the subgroups and identify their elements.

6. In Table 15.21, we presented a 3×3 matrix representation of the D_3 group.

   (a) Find the transformation matrix that puts this representation in block diagonal form.

   (b) What is the inverse of this transformation matrix?

   (c) Show that the character of the matrices in the block diagonal form are the same as the original matrices.

7. Consider the twofold symmetric molecule composed of three pairs of atoms shown below. In the chapter, we developed the D_2 group to represent the symmetry of this object. The four elements of this group satisfy the vierergruppe multiplication table given in Table 15.7.
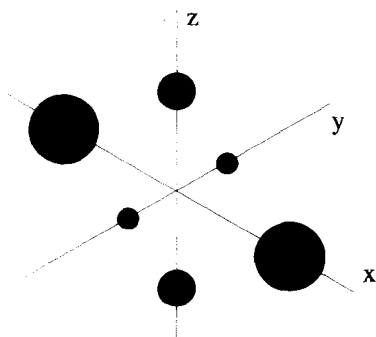
The symmetry of this molecule allows the identification of a group of larger order. For example, the interchange of only the two largest atoms will also leave the structure unchanged and is not represented by an element of D_2. The D_2 group describes only rigid body transformations. Interchanging the two largest atoms while leaving the others alone is not a rigid body transformation.

   (a) Identify all the elements of this higher-order group and determine its multiplication table.

   (b) Show that D_2 is a subgroup of this group.

   (c) Develop a 3×3 matrix representation for this group.

   (d) Using the [123456] operator representation, identify the elements of this higher-order group.

   (e) Verify that h/h' is an integer, where h is the order of the larger group and h' is the order of D_2.

   (f) Determine the left cosets of the D_2 subgroup.

8. Consider the symmetry group of the regular tetrahedron shown below. If nonrigid body transformations are allowed, this becomes the S_4 group.
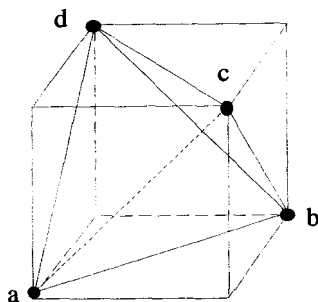
(a) What is the order of S_4?

(b) Obtain a nontrivial 4×4 matrix representation for this group.

(c) Identify the classes of this group and show that the character of each of the matrices in part (b) is the same for each class.

(d) Identify the subgroup of S_4 that corresponds to only the rigid rotations of the regular tetrahedron.

9. Construct a 4×4 matrix representation for the C_4 group. The construction of this representation, which we will call $C_4^{[X]}$, can be accomplished using the following method. According to the C_4 multiplication table, $g_3\, g_2 = g_4$. We will require that the $g_3^{[X]}$ element obey

$$g_3^{[X]} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Likewise, $g_3\, g_3 = I$, so we will also require

$$g_3^{[X]} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

(a) Continue this process and determine all four elements of the $C_4^{[X]}$ representation.

(b) Determine the characters of the $C_4^{[X]}$ representation. Compare them to the characters of the irreducible representations of C_4.

(c) The $C_4^{[X]}$ representation is reducible. Using the completeness and orthogonality of the character table of C_4, determine the number of times each irreducible representation of C_4 appears in the reduced form of $C_4^{[X]}$.

(d) Give the transformation matrix that puts the representation into the reduced form in part (c).

10. Consider an object consisting of four identical atoms, as shown below:
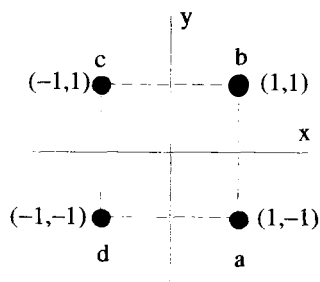
Let the D_4 group consist of the set of rigid body transformations that leave the appearance of this object unchanged.

(a) Identify all the elements of this group using the [1234] operator notation.

(b) Construct the multiplication table for D_4.

(c) Identify the inverse of each element of the group.

(d) Identify all the subgroups of D_4.

(e) Without performing any similarity transforms, identify the different classes associated with D_4.

(f) How many possible irreducible matrix representations are there for this group?

(g) A 2×2 matrix representation can be constructed for this group by associating a two-dimensional transformation matrix with each element. Determine all the elements of this 2×2 representation.

If the restriction of rigid body transformations is relaxed, the symmetry of this object can be described by the S_4 symmetry group, which consists of all the permutations of four objects.

(h) Identify all the elements of S_4 using [1234] operator notation.

(i) Show that D_4 is a subgroup of S_4.

11. Construct a 6×6 matrix representation for the D_3 group, following the same prescription used in Exercise 9. We will call this representation $D_3^{[X]}$.

(a) Determine the character of the elements of the $D_3^{[X]}$ representations and include them as a fourth column in the character table for the D_3 group, as shown below:

	$\chi^{[1]}$	$\chi^{[2]}$	$\chi^{[3]}$	$\chi^{[X]}$
I	1	1	2	?
g_2	1	1	-1	?
g_3	1	1	-1	?
g_4	1	-1	0	?
g_5	1	-1	0	?
g_6	1	-1	0	?

(b) Show that the vectors formed by the character values of the irreducible representations are all orthogonal to each other, but not to the vector formed by the character values of the $D_3^{[X]}$ representation.

(c) Determine how many times each of the irreducible representations appears in the decomposition of the $D_3^{[X]}$ representation.

12. Consider the symmetry group associated with a diatomic molecule. This molecule has two identical atoms located in the xy-plane, as shown below:
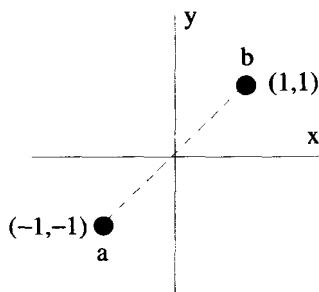
(a) Determine the four elements of the symmetry group, which correspond to rigid rotations in the plane and reflections.

(b) Construct a 2×2 matrix representation of this group by associating a transformation matrix with each element.

(c) Determine the multiplication table of this group.

(d) Find the transformation that simultaneously diagonalizes all the elements of the representation found for part (b) and list all the diagonalized matrix elements.

(e) The transformation you found in part (d) can be viewed as a coordinate system transformation. Sketch the locations of the atoms in this new coordinate system.

13. In Equation 15.68, we derived a matrix representation for the g_2 element of D_3. Derive the other elements of this representation, which are listed in Table 15.30.

14. Another way to write the general form of the continuous $SU(2)$ group is

$$\begin{bmatrix} e^{i\alpha} \cos \gamma & e^{i\beta} \sin \gamma \\ -e^{-i\beta} \sin \gamma & e^{-i\alpha} \cos \gamma \end{bmatrix}, \qquad (15.89)$$

where α, β, and γ are all real quantities. Show that this form automatically satisfies both the unimodular and unitary conditions.

A

THE LEVI-CIVITA IDENTITY

The three-dimensional Levi-Civita symbol is defined as

$$\epsilon_{ijk} = \begin{cases} +1 & \text{for } i, j, k = \text{even permutations of } 1, 2, 3 \\ -1 & \text{for } i, j, k = \text{odd permutations of } 1, 2, 3 \\ 0 & \text{if two or more of the subscripts are equal} \end{cases} \qquad (A.1)$$

One useful identity associated with this symbol is

$$\epsilon_{ijk}\epsilon_{rsk} = \delta_{ir}\delta_{js} - \delta_{is}\delta_{jr}. \qquad (A.2)$$

To prove Equation A.2, start by explicitly writing out the sum over k on the LHS:

$$\epsilon_{ijk}\epsilon_{rsk} = \epsilon_{ij1}\epsilon_{rs1} + \epsilon_{ij2}\epsilon_{rs2} + \epsilon_{ij3}\epsilon_{rs3}. \qquad (A.3)$$

Now consider $\epsilon_{ij1}\epsilon_{rs1}$, the first term on the RHS of Equation A.3. This term is zero if i, j, r, or s is equal to 1. It is also zero if $i = j$ or $r = s$. Therefore, $\epsilon_{ij1}\epsilon_{rs1}$ is nonzero for only four combinations of the indices (i, j, r, s): $(2, 3, 2, 3)$, $(2, 3, 3, 2)$, $(3, 2, 2, 3)$, and $(3, 2, 3, 2)$. Thus the first term on the RHS of Equation A.3 becomes

$$\epsilon_{ij1}\epsilon_{rs1} = \begin{cases} +1 & \text{for } (i, j, r, s) = (2, 3, 2, 3) \text{ or } (3, 2, 3, 2) \\ -1 & \text{for } (i, j, r, s) = (2, 3, 3, 2) \text{ or } (3, 2, 2, 3) \\ 0 & \text{otherwise} \end{cases} \qquad (A.4)$$

The other two terms on the RHS of Equation A.3 behave in a similar fashion, with the results:

$$\epsilon_{ij2}\epsilon_{rs2} = \begin{cases} +1 & \text{for } (i, j, r, s) = (1, 3, 1, 3) \text{ or } (3, 1, 3, 1) \\ -1 & \text{for } (i, j, r, s) = (1, 3, 3, 1) \text{ or } (3, 1, 1, 3) \\ 0 & \text{otherwise} \end{cases} \qquad (A.5)$$

and

$$\epsilon_{ij3}\epsilon_{rs3} = \begin{cases} +1 & \text{for } (i,j,r,s) = (1,2,1,2) \text{ or } (2,1,2,1) \\ -1 & \text{for } (i,j,r,s) = (1,2,2,1) \text{ or } (2,1,1,2) \ . \\ 0 & \text{otherwise} \end{cases} \tag{A.6}$$

The LHS of Equation A.2 is the sum of the three terms described above. For a given set of values for $i, j, r,$ and s with $i \neq j$ and $r \neq s$, only one of the terms is nonzero. Therefore,

$$\epsilon_{ijk}\epsilon_{rsk} = \begin{cases} +1 & \text{for } i \neq j, r \neq s, i = r, j = s \\ -1 & \text{for } i \neq j, r \neq s, i = s, j = r \ . \\ 0 & \text{otherwise} \end{cases} \tag{A.7}$$

But notice, these conditions are also satisfied by the quantity $\delta_{ir}\delta_{js} - \delta_{is}\delta_{jr}$, because

$$\delta_{ir}\delta_{js} - \delta_{is}\delta_{jr} = \begin{cases} +1 & \text{for } i \neq j, r \neq s, j = s \\ -1 & \text{for } i \neq j, r \neq s, j = r \ . \\ 0 & \text{otherwise} \end{cases} \tag{A.8}$$

Hence, we have our result:

$$\epsilon_{ijk}\epsilon_{rsk} = \delta_{ir}\delta_{js} - \delta_{is}\delta_{jr}. \tag{A.9}$$

B

THE CURVILINEAR CURL

A rigorous derivation of the form of the curl, gradient, and divergence operators in curvilinear coordinate systems requires a careful analysis of the scale factors and how they change as functions of position. In the text, this was treated in a somewhat cavalier manner. In this appendix, we present a more rigorous derivation of the curvilinear curl operation. The final result is, of course, the same as presented earlier. The rigorous derivation of the gradient and divergence operations follow a similar manner.

The curl operation can be defined by the integral relation

$$\int_S d\overline{\boldsymbol{\sigma}} \cdot \overline{\boldsymbol{\nabla}} \times \overline{\mathbf{A}} = \oint_C d\overline{\mathbf{r}} \cdot \overline{\mathbf{A}} \tag{B.1}$$

where C is a closed path surrounding the surface S, and the direction of $d\overline{\boldsymbol{\sigma}}$ is defined by the direction of C and the right-hand rule. If we align the surface perpendicular to the 1-direction, as shown in Figure B.1, and take the limit as the surface shrinks to a differential area, the LHS becomes

$$\lim_{S \to 0} \int_S d\overline{\boldsymbol{\sigma}} \cdot \overline{\boldsymbol{\nabla}} \times \overline{\mathbf{A}} = \underbrace{\hat{q}_1 \cdot \left[\overline{\boldsymbol{\nabla}} \times \overline{A}\right]}_{[\overline{\nabla} \times \overline{A}]_1} h_2 h_3 dq_2 dq_3. \tag{B.2}$$

This term is proportional to the 1-component of the curl.

The RHS of Equation B.1 in this differential limit is

$$\lim_{C \to 0} \oint_C d\overline{\mathbf{r}} \cdot \overline{\mathbf{A}} = \int_{C_1} d\overline{\mathbf{r}} \cdot \overline{\mathbf{A}} + \int_{C_2} d\overline{\mathbf{r}} \cdot \overline{\mathbf{A}} + \int_{C_3} d\overline{\mathbf{r}} \cdot \overline{\mathbf{A}} + \int_{C_4} d\overline{\mathbf{r}} \cdot \overline{\mathbf{A}}, \tag{B.3}$$
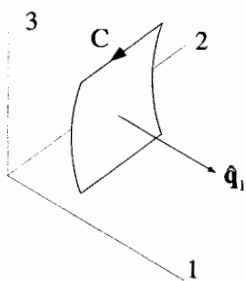
Figure B.1   Orientation of Surface for Curvilinear Curl Integration

where the loop has been broken up into four parts, as shown in Figure B.2. The contribution from C_1 is given by

$$\int_{C_1} d\bar{\mathbf{r}} \cdot \bar{\mathbf{A}} = \int_{q_{2o}}^{q_{2o}+dq_2} dq_2 A_2 h_2. \tag{B.4}$$

Along C_1, both A_2 and h_2 can vary with q_2. Since we are in the differential limit, we will consider only the lowest-order, linear variation of the integrand:

$$A_2 h_2 = (A_2 h_2)\Big|_{q_{2o},q_{3o}} + \frac{\partial(A_2 h_2)}{\partial q_2}\Big|_{q_{2o},q_{3o}} (q_2 - q_{2o}). \tag{B.5}$$

Substituting Equation B.5 into Equation B.4 gives

$$\int_{C_1} d\bar{\mathbf{r}} \cdot \bar{\mathbf{A}} = (A_2 h_2)\Big|_{q_{2o},q_{3o}} dq_2 + \frac{\partial(A_2 h_2)}{\partial q_2}\Big|_{q_{2o},q_{3o}} \frac{(dq_2)^2}{2}. \tag{B.6}$$

Next the integration along C_3 will be performed:

$$\int_{C_3} d\bar{\mathbf{r}} \cdot \bar{\mathbf{A}} = \int_{q_{2o}+dq_2}^{q_{2o}} dq_2 A_2 h_2. \tag{B.7}$$
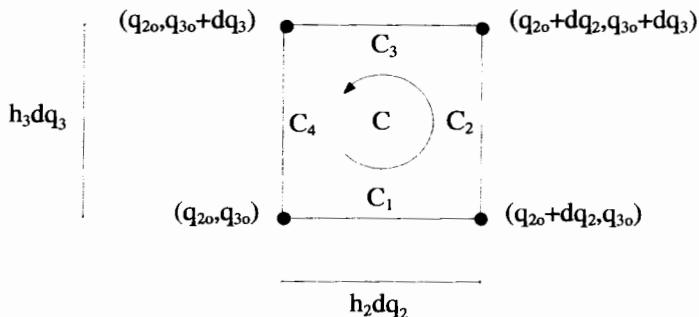
Figure B.2   Differential Geometry for Curvilinear Curl Integrations

Here the expression for $A_2 h_2$ picks up an extra term because $q_3 = q_{3o} + dq_3$:

$$A_2 h_2 = (A_2 h_2)\Big|_{q_{2o}, q_{3o}} + \frac{\partial (A_2 h_2)}{\partial q_2}\Big|_{q_{2o}, q_{3o}} (q_2 - q_{2o}) + \frac{\partial (A_2 h_2)}{\partial q_3}\Big|_{q_{2o}, q_{3o}} dq_3. \quad \text{(B.8)}$$

Substituting Equation B.8 into Equation B.7 gives for the line integral along C_3,

$$\int_{C_3} d\bar{\mathbf{r}} \cdot \overline{\mathbf{A}} = -(A_2 h_2)\Big|_{q_{2o}, q_{3o}} dq_2 - \frac{\partial (A_2 h_2)}{\partial q_2}\Big|_{q_{2o}, q_{3o}} \frac{(dq_2)^2}{2}$$

$$- \frac{\partial (A_2 h_2)}{\partial q_3}\Big|_{q_{2o}, q_{3o}} dq_2 dq_3. \quad \text{(B.9)}$$

Similar integrations can be performed along C_2 and C_4. When summed, these pieces give the value for the closed line integral:

$$\lim_{C \to 0} \oint_C d\bar{\mathbf{r}} \cdot \overline{\mathbf{A}} = \left[\frac{\partial (h_3 A_3)}{\partial q_2} - \frac{\partial (h_2 A_2)}{\partial q_3} \right] dq_2 dq_3. \quad \text{(B.10)}$$

Notice how the higher-order terms have all canceled. Substituting Equations B.10 and B.2 into Equation B.1 gives for the 1-component of the curl of $\overline{\mathbf{A}}$,

$$\left[\boldsymbol{\nabla} \times \overline{\mathbf{A}} \right]_1 = \frac{1}{h_2 h_3} \left[\frac{\partial (h_3 A_3)}{\partial q_2} - \frac{\partial (h_2 A_2)}{\partial q_3} \right]. \quad \text{(B.11)}$$

The other components of $\overline{\boldsymbol{\nabla}} \times \overline{\mathbf{A}}$ can be obtained by reorientating the surface shown in Figure B.1 to be along the other curvilinear axes:

$$\left[\boldsymbol{\nabla} \times \overline{\mathbf{A}} \right]_2 = \frac{1}{h_1 h_3} \left[\frac{\partial (h_1 A_1)}{\partial q_3} - \frac{\partial (h_3 A_3)}{\partial q_1} \right] \quad \text{(B.12)}$$

$$\left[\boldsymbol{\nabla} \times \overline{\mathbf{A}} \right]_3 = \frac{1}{h_1 h_2} \left[\frac{\partial (h_2 A_2)}{\partial q_1} - \frac{\partial (h_1 A_1)}{\partial q_2} \right]. \quad \text{(B.13)}$$

Equations B.11–B.13 can be written compactly using a determinant:

$$\boldsymbol{\nabla} \times \overline{\mathbf{A}} = \frac{1}{h_1 h_2 h_3} \begin{vmatrix} h_1 \hat{\mathbf{q}}_1 & h_2 \hat{\mathbf{q}}_2 & h_3 \hat{\mathbf{q}}_3 \\ \partial / \partial q_1 & \partial / \partial q_2 & \partial / \partial q_3 \\ h_1 A_1 & h_2 A_2 & h_3 A_3 \end{vmatrix}_{det}, \quad \text{(B.14)}$$

or even more concisely using the Levi-Civita symbol and subscript/summation notation:

$$\boldsymbol{\nabla} \times \overline{\mathbf{A}} = \epsilon_{ijk} \frac{\hat{\mathbf{q}}_i}{h_j h_k} \frac{\partial (h_k A_k)}{\partial q_j}. \quad \text{(B.15)}$$

This Page Intentionally Left Blank

C

THE DOUBLE INTEGRAL IDENTITY

In this appendix, the double integral identity

$$\int_a^x dx'' \int_a^{x''} dx'\, f(x') = \int_a^x dx'\, (x - x')f(x') \tag{C.1}$$

is proven. To derive this relation, make the substitution

$$f(x) = \frac{dy(x)}{dx}. \tag{C.2}$$

Equation C.1 becomes

$$\int_a^x dx'' \int_a^{x''} dx'\frac{dy(x')}{dx'} = \int_a^x dx'(x - x')\frac{dy(x')}{dx'}. \tag{C.3}$$

The first integration on the LHS, over the variable x', can be done directly. The integral on the RHS can be rewritten using integration by parts, with $u = (x - x')$ and $dv = (dy/dx')dx'$, to give

$$\int_a^x dx''[y(x'') - y(a)] = (x - x')y(x')\Big|_{x'=a}^{x'=x} + \int_a^x dx'\, y(x'). \tag{C.4}$$

When integration of $y(a)$ on the LHS is performed, and the first term on the RHS is evaluated at the limits, the result is

$$(a - x)y(a) + \int_a^x dx''\, y(x'') = (a - x)y(a) + \int_a^x dx'\, y(x'). \tag{C.5}$$

Since x' and x'' are dummy variables of integration, the two sides of this equation are obviously equal, and the identity of Equation C.1 is proven.

This Page Intentionally Left Blank

D

GREEN'S FUNCTION SOLUTIONS

In Chapter 10, the Green's function solution for a linear, nonhomogeneous, second-order equation was obtained using a superposition argument. We first solved the special case of a δ-function drive for the nonhomogeneous term. Then we argued that the general solution, with an arbitrary nonhomogeneous term, could be obtained by simply forming an infinite weighted sum of these δ-function solutions. This proof relied on a certain amount of intuition. A formal proof of this solution can be obtained by directly inserting the Green's function solution into the differential equation, and showing that the nonhomogeneous differential equation and the boundary conditions are satisfied.

Consider the general linear, nonhomogeneous, second-order differential equation:

$$\frac{d}{dx}\left[P(x)\frac{dy(x)}{dx}\right] - Q(x)y(x) = h(x). \tag{D.1}$$

This equation is in Sturm-Liouville form, but keep in mind any linear, second-order equation can be put into this form with a little manipulation. We want to find the function $y(x)$ which solves Equation D.1 and satisfies the set of homogeneous boundary conditions:

$$\alpha_1 y(a) + \alpha_2 \frac{dy(x)}{dx}\bigg|_{x=a} = 0 \tag{D.2}$$

$$\beta_1 y(b) + \beta_2 \frac{dy(x)}{dx}\bigg|_{x=b} = 0. \tag{D.3}$$

We will prove that the integral

$$y(x) = \int_a^b d\xi \, h(\xi)g(x|\xi) \tag{D.4}$$

satisfies these conditions, where $g(x|\xi)$ is a function that satisfies the differential equation

$$\frac{d}{dx}\left[P(x)\frac{dg(x|\xi)}{dx}\right] - Q(x)g(x|\xi) = \delta(x - \xi) \tag{D.5}$$

and the same set of homogeneous boundary conditions:

$$\alpha_1 g(a|\xi) + \alpha_2 \frac{dg(x|\xi)}{dx}\bigg|_{x=a} = 0 \tag{D.6}$$

$$\beta_1 g(b|\xi) + \beta_2 \frac{dg(x|\xi)}{dx}\bigg|_{x=b} = 0. \tag{D.7}$$

We begin by describing the properties of the Green's function $g(x|\xi)$. Notice everywhere, except at $x = \xi$, $g(x|\xi)$ satisfies the homogeneous differential equation

$$\frac{d}{dx}\left[P(x)\frac{dg(x|\xi)}{dx}\right] - Q(x)g(x|\xi) = 0. \tag{D.8}$$

If $g_1(x|\xi)$ and $g_2(x|\xi)$ are the solutions to this homogeneous equation on either side of $x = \xi$, we can write

$$g(x|\xi) = \begin{cases} g_1(x|\xi) & a \le x \le \xi \\ g_2(x|\xi) & \xi \le x \le b \end{cases} . \tag{D.9}$$

Since Equation D.8 is a second-order equation, its general solution will contain two arbitrary constants. But these constants are not necessarily the same on the two sides of $x = \xi$, so there really are a total of four unknown constants associated with Equation D.9, two for $g_1(x|\xi)$ and two for $g_2(x|\xi)$. Therefore, we need four independent equations to completely specify $g(x|\xi)$. Two come directly from the boundary conditions. Using $g_1(x|\xi)$ in Equation D.6 establishes one equation. Equation D.7 with $g_2(x|\xi)$ establishes the second. A third equation comes from requiring that $g(x|\xi)$ be continuous at $x = \xi$:

$$g_1(\xi|\xi) = g_2(\xi|\xi). \tag{D.10}$$

This continuity is guaranteed by the form of Equation D.5. If $g(x|\xi)$ were discontinuous at any point, the second derivative operation on the LHS of Equation D.5 would generate a term proportional to the derivative of a δ-function, a term that obviously does not exist on the RHS of Equation D.5. The last equation comes from integration

of Equation D.5 across the δ-function:

$$\int_{\xi-\epsilon}^{\xi+\epsilon} dx \left\{ \frac{d}{dx} \left[P(x) \frac{dg(x|\xi)}{dx} \right] - Q(x)g(x|\xi) \right\} = \int_{\xi-\epsilon}^{\xi+\epsilon} dx \, \delta(x - \xi). \quad \text{(D.11)}$$

For arbitrarily small ϵ this becomes

$$P(\xi) \frac{dg(x|\xi)}{dx} \Bigg|_{x=\xi-\epsilon}^{x=\xi+\epsilon} - Q(\xi)g(\xi|\xi)2\epsilon = 1, \quad \text{(D.12)}$$

where we have assumed that $Q(x)$ and $P(x)$ are well-behaved functions, so near $x = \xi$, $Q(x) \rightarrow Q(\xi)$ and $P(x) \rightarrow P(\xi)$. As $\epsilon \rightarrow 0$, Equation D.12 becomes

$$\frac{dg_2(x|\xi)}{dx} \Bigg|_{x=\xi} - \frac{dg_1(x|\xi)}{dx} \Bigg|_{x=\xi} = \frac{1}{P(\xi)}. \quad \text{(D.13)}$$

The δ-function on the RHS of Equation D.5 causes a discontinuity in the derivative of $g(x|\xi)$ at $x = \xi$.

Next we show that Equation D.4 is a solution to the original differential equation, by substituting it into Equation D.1. In terms of $g_1(x|\xi)$ and $g_2(x|\xi)$, this solution is

$$y(x) = \int_a^x d\xi \, h(\xi)g_2(x|\xi) + \int_x^b d\xi \, h(\xi)g_1(x|\xi). \quad \text{(D.14)}$$

Before putting this in the LHS of Equation D.1, we need to evaluate its first and second derivatives. This is a bit tricky, because the derivatives are taken with respect to a variable that is in both the integrand and the limits of integration. To evaluate such a derivative, use the result

$$\frac{d}{dx} \int_{a(x)}^{b(x)} dy \, C(x, y) = \int_{a(x)}^{b(x)} dy \, \frac{\partial C(x, y)}{\partial x} + \frac{db(x)}{dx} C(x, b(x)) - \frac{da(x)}{dx} C(x, a(x)).$$

$$\text{(D.15)}$$

Therefore,

$$\frac{dy(x)}{dx} = \int_a^x d\xi \, h(\xi) \frac{dg_2(x|\xi)}{dx} + \int_x^b d\xi \, h(\xi) \frac{dg_1(x|\xi)}{dx} + h(x)g_2(x|x) - h(x)g_1(x|x).$$

$$\text{(D.16)}$$

The last two terms in this expression cancel because the Green's function is continuous, and we have

$$\frac{dy(x)}{dx} = \int_a^x d\xi \, h(\xi) \frac{dg_2(x|\xi)}{dx} + \int_x^b d\xi \, h(\xi) \frac{dg_1(x|\xi)}{dx}. \quad \text{(D.17)}$$

The second derivative of $y(x)$ is

$$\frac{d^2y(x)}{dx^2} = \int_a^x d\xi\, h(\xi)\frac{d^2g_2(x|\xi)}{dx^2} + \int_x^b d\xi\, h(\xi)\frac{d^2g_1(x|\xi)}{dx^2}$$

$$+ h(x)\frac{dg_2(x|\xi)}{dx}\bigg|_{x=\xi} - h(x)\frac{dg_1(x|\xi)}{dx}\bigg|_{x=\xi}. \qquad (D.18)$$

The last two terms of this expression do not cancel because of the discontinuity in the derivative of the Green's function we discovered earlier. Using Equation D.13, however, these terms can be combined to give

$$\frac{d^2y(x)}{dx^2} = \int_a^x d\xi\, h(\xi)\frac{d^2g_2(x|\xi)}{dx^2} + \int_x^b d\xi\, h(\xi)\frac{d^2g_1(x|\xi)}{dx^2} + \frac{h(x)}{P(x)}. \qquad (D.19)$$

Substituting these expressions for $y(x)$ and its derivatives into the LHS of Equation D.1 gives

$$P(x)\left[\int_a^x d\xi\, h(\xi)\frac{d^2g_2(x|\xi)}{dx^2} + \int_x^b d\xi\, h(\xi)\frac{d^2g_1(x|\xi)}{dx^2} + \frac{h(x)}{P(x)}\right]$$

$$+ \frac{dP(x)}{dx}\left[\int_a^x d\xi\, h(\xi)\frac{dg_2(x|\xi)}{dx} + \int_x^b d\xi h(\xi)\frac{dg_1(x|\xi)}{dx}\right]$$

$$- Q(x)\left[\int_a^x d\xi\, h(\xi)g_2(x|\xi) + \int_x^b d\xi\, h(\xi)g_1(x|\xi)\right]. \qquad (D.20)$$

Both $g_1(x|\xi)$ and $g_2(x|\xi)$ are solutions to the homogeneous Equation D.8. Because of this, all the terms in Equation D.20 cancel, except the single term $h(x)$. Our final result is

$$\left\{\frac{d}{dx}\left[P(x)\frac{d}{dx}\right] - Q(x)\right\}\int_a^b d\xi\, h(\xi)g(x|\xi) = h(x), \qquad (D.21)$$

which shows that the Green's function solution for $y(x)$ indeed satisfies the original nonhomogeneous differential equation.

It also easy to show that the Green's function solution satisfies the original homogeneous boundary conditions. The boundary condition at $x = a$ is given by Equation D.2, and using Equation D.14 and D.17 evaluated at $x = a$, we can write

$$\alpha_1 y(a) + \alpha_2\frac{dy(x)}{dx}\bigg|_{x=a} = \alpha_1\int_a^b d\xi\, h(\xi)g_1(a|\xi) + \alpha_2\int_a^b d\xi\, h(\xi)\frac{dg_1(x|\xi)}{dx}\bigg|_{x=a}$$

$$= \int_a^b d\xi\, h(\xi)\left[\alpha_1 g_1(a|\xi) + \alpha_2\frac{dg_1(x|\xi)}{dx}\bigg|_{x=a}\right]$$

$$= 0. \qquad (D.22)$$

Since we constructed the Green's function to obey the same boundary condition, as given by Equation D.6, the last line in Equation D.22 is zero. A similar line of reasoning shows that the boundary condition at $x = b$, given by Equation D.7, also works.

This Page Intentionally Left Blank

E

PSEUDOVECTORS AND
THE MIRROR TEST

The mirror test provides an intuitive way to identify a pseudovector. A "regular" polar vector will reflect, as you would expect, in a mirror but a pseudovector will be reversed.

An example of this test, taken from electromagnetism, is described in Figure E.1. The real system on the left is a positively charged particle moving in a circular orbit, counterclockwise when viewed from above. This motion results in the average
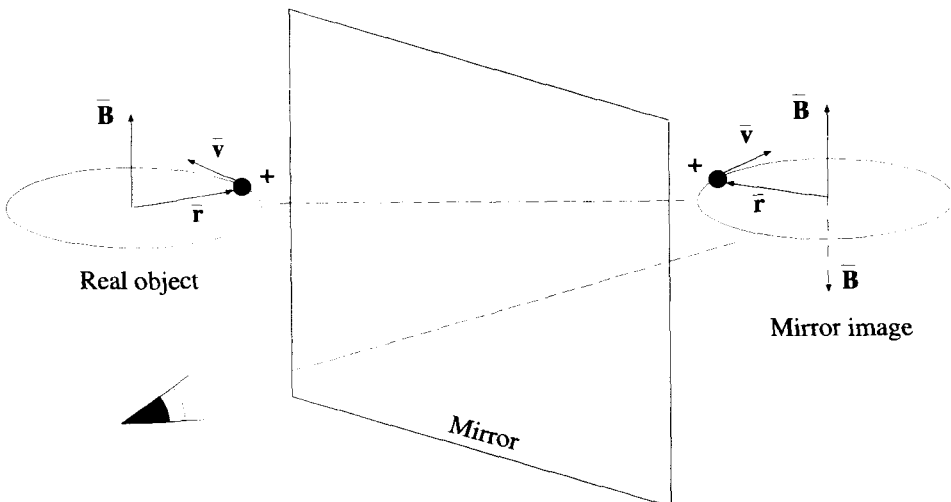
Figure E.1   The Mirror Test for Pseudovectors

magnetic field pointing "up," as indicated by the $\overline{\mathbf{B}}$ vector. At a particular instant of time, this particle is at a position indicated by $\overline{r}$ and moving at a velocity $\overline{v}$. Now view the image of this motion in the mirror shown in Figure E.1. The image particle moves in a clockwise, circular orbit. This motion is properly described by the images of the position and velocity vectors. Hence, they are regular vectors. The reflected image of the magnetic field vector is still in the up direction. But notice, if a positively charged particle were really performing the motion of this image particle, the magnetic field would point down, as indicated by the dashed $\overline{\mathbf{B}}$ vector. The magnetic field vector fails the mirror test, and is therefore a pseudovector.

F

CHRISTOFFEL SYMBOLS AND COVARIANT DERIVATIVES

Non-orthonormal coordinate systems become more complicated if the basis vectors are position dependent. An example of a two-dimensional coordinate system of this type is shown in Figure F.1. In this system, the displacement vector can still be written as

$$d\bar{\mathbf{r}} = dx^1\,\hat{\mathbf{g}}_1 + dx^2\,\hat{\mathbf{g}}_2. \tag{F.1}$$

Since

$$d\bar{\mathbf{r}} = \left(\frac{\partial\bar{\mathbf{r}}}{\partial x^1}\right)dx^1 + \left(\frac{\partial\bar{\mathbf{r}}}{\partial x^2}\right)dx^2, \tag{F.2}$$

the covariant basis vectors are still identified as

$$\hat{\mathbf{g}}_i = \frac{\partial\bar{\mathbf{r}}}{\partial x^i}, \tag{F.3}$$

but now are functions of position. The effect of nonconstant basis vectors is most evident when applying derivatives to vector and scalar fields.

In Chapter 14, the gradient operation was discussed for skewed coordinate systems, where the basis vectors were not orthonormal and had constant basis vectors. The gradient was defined as

$$\overline{\boldsymbol{\nabla}}\phi = \hat{\mathbf{g}}^i\frac{\partial\phi}{\partial x^i}, \tag{F.4}$$
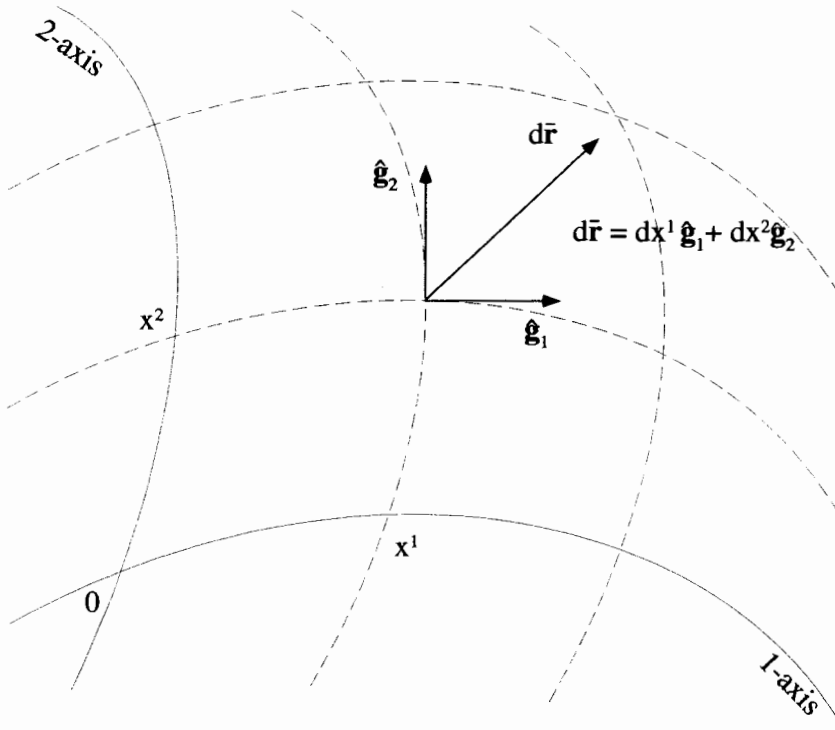
Figure F.1 A Coordinate System in Curved Space

and the del operator itself was identified as

$$\overline{\nabla} \equiv \hat{\mathbf{g}}^i \frac{\partial}{\partial x^i}. \tag{F.5}$$

The derivative $\partial\phi/\partial x^i$, is the ith covariant component of the gradient vector. Notice how the contravariant basis vector $\hat{\mathbf{g}}^i$ is not differentiated. This is important, because when we move to systems where the basis vectors are no longer constants, the gradient operation does not need to be modified in any way.

Things are very different for the divergence of a vector field. Using the same definition for the del operator, the divergence of a vector $\overline{\mathbf{V}} = V^i \hat{\mathbf{g}}_i$ becomes

$$\overline{\nabla} \cdot \overline{\mathbf{V}} = \hat{\mathbf{g}}^j \frac{\partial}{\partial x^j} \cdot \left(V^i \hat{\mathbf{g}}_i \right). \tag{F.6}$$

Now the spatial derivative operates both on the vector components V^i and the $\hat{\mathbf{g}}_i$ basis vectors. Rearranging Equation F.6 gives

$$\overline{\nabla} \cdot \overline{\mathbf{V}} = \hat{\mathbf{g}}^j \cdot \frac{\partial}{\partial x^j} \left(V^i \hat{\mathbf{g}}_i \right). \tag{F.7}$$

If the basis vectors are not constants, the RHS of Equation F.7 generates two terms

$$\overline{\nabla} \cdot \overline{\mathbf{V}} = \hat{\mathbf{g}}^j \cdot \left[\left(\frac{\partial V^i}{\partial x^j} \right) \hat{\mathbf{g}}_i + V^i \left(\frac{\partial \hat{\mathbf{g}}_i}{\partial x^j} \right) \right]. \tag{F.8}$$

The last term in Equation F.8 is usually defined in terms of the *Christoffel symbol* Γ^k_{ij}:

$$\frac{\partial \hat{\mathbf{g}}_i}{\partial x^j} = \Gamma^k_{ij} \hat{\mathbf{g}}_k. \tag{F.9}$$

The definition in Equation F.9 implies the result of the differentiation on the LHS must be a vector quantity, expressed in terms of the covariant basis vectors $\hat{\mathbf{g}}_k$. The contravariant components of the vector quantity are given by the Christoffel symbol with a superscripted k. These components obviously also depend on which basis vector is being differentiated, given by the i index, and which coordinate the differentiation is being taken with respect to, given by the j index. Because of the nature of the LHS of Equation F.9, these indices are written as subscripts, so that the Christoffel symbol in Equation F.9 obeys the normal superscript/subscript conventions. It is important to note, however, the Christoffel symbol is not a tensor. Its elements do not transform like the elements of a tensor.

In order to complete our discussion of the divergence, we must evaluate the Christoffel symbols in terms of the coordinate system geometry. Equation F.9 can be solved for Γ^k_{ij} by dot multiplying both sides by $\hat{\mathbf{g}}^l$:

$$\hat{\mathbf{g}}^l \cdot \frac{\partial \hat{\mathbf{g}}_i}{\partial x^j} = \Gamma^k_{ij} (\hat{\mathbf{g}}^l \cdot \hat{\mathbf{g}}_k). \tag{F.10}$$

or

$$\Gamma^l_{ij} = \hat{\mathbf{g}}^l \cdot \frac{\partial \hat{\mathbf{g}}_i}{\partial x^j}. \tag{F.11}$$

The basis vectors can still be written in terms of derivatives of the position vector,

$$\hat{\mathbf{g}}_i = \frac{\partial \overline{\mathbf{r}}}{\partial x^i}, \tag{F.12}$$

so the Christoffel symbol becomes

$$\Gamma^l_{ij} = \hat{\mathbf{g}}^l \cdot \frac{\partial^2 \overline{\mathbf{r}}}{\partial x^j \partial x^i}. \tag{F.13}$$

This equation clearly indicates that the Christoffel symbol has a symmetry with respect to the subscripted indices

$$\Gamma^k_{ij} = \Gamma^k_{ji}. \tag{F.14}$$

Equation F.13 provides a method for evaluating the Christoffel symbol, but it is not very useful because it is in terms of the derivatives of the position vector and the basis vectors of the coordinate system. A more useful relation can be derived by

considering the metric. Remember the metric for a coordinate system is

$$M_{ij} = \hat{\mathbf{g}}_i \cdot \hat{\mathbf{g}}_j. \tag{F.15}$$

Even though the Christoffel symbol is not a tensor, this metric can be used to define a new set of quantities:

$$\Gamma_{kij} \equiv M_{kl}\Gamma^l_{ij}. \tag{F.16}$$

This quantity, Γ_{kij}, is often called a Christoffel symbol of the first kind, while Γ^k_{ij} is a Christoffel symbol of the second kind. Notice the Christoffel symbol of the first kind exhibits the same symmetry with respect to the last two subscripts:

$$\Gamma_{kij} = \Gamma_{kji}. \tag{F.17}$$

Combining Equations F.11 and F.16 gives

$$\Gamma_{kij} = M_{kl}\hat{\mathbf{g}}^l \cdot \frac{\partial \hat{\mathbf{g}}_i}{\partial x^j}$$

$$= \hat{\mathbf{g}}_k \cdot \frac{\partial \hat{\mathbf{g}}_i}{\partial x^j}. \tag{F.18}$$

The spatial derivative of the metric,

$$\frac{\partial M_{ij}}{\partial x^k} = \hat{\mathbf{g}}_i \cdot \left(\frac{\partial \hat{\mathbf{g}}_j}{\partial x^k}\right) + \left(\frac{\partial \hat{\mathbf{g}}_i}{\partial x^k}\right) \cdot \hat{\mathbf{g}}_j, \tag{F.19}$$

can now be written in terms of Christoffel symbols as

$$\frac{\partial M_{ij}}{\partial x^k} = \Gamma_{ijk} + \Gamma_{jik}. \tag{F.20}$$

Similarly,

$$\frac{\partial M_{ik}}{\partial x^j} = \Gamma_{ikj} + \Gamma_{kij} \tag{F.21}$$

and

$$\frac{\partial M_{jk}}{\partial x^i} = \Gamma_{jki} + \Gamma_{kji}. \tag{F.22}$$

Adding Equation F.22 to F.21 and subtracting F.20, while making use of the symmetry of the Christoffel symbols, gives

$$\Gamma_{kij} = \frac{1}{2}\left(\frac{\partial M_{ik}}{\partial x^j} + \frac{\partial M_{jk}}{\partial x^i} - \frac{\partial M_{ij}}{\partial x^k}\right). \tag{F.23}$$

Raising the first index with the metric gives an expression for the Christoffel symbol of the second kind in terms of the coordinate system's metric:

$$\Gamma^k_{\ ij} = \frac{M^{kl}}{2} \left(\frac{\partial M_{il}}{\partial x^j} + \frac{\partial M_{jl}}{\partial x^i} - \frac{\partial M_{ij}}{\partial x^l} \right). \tag{F.24}$$

This equation allows us to evaluate the Christoffel symbol if we know the metric.

Returning to the divergence operation, Equation F.8 can now be written using the Christoffel symbol as

$$\overline{\nabla} \cdot \overline{V} = \hat{\mathbf{g}}^j \cdot \frac{\partial}{\partial x^j} \left(V^i \hat{\mathbf{g}}_i \right)$$

$$= \hat{\mathbf{g}}^j \cdot \left[\left(\frac{\partial V^i}{\partial x^j} \hat{\mathbf{g}}_i \right) + V^i \Gamma^k_{\ ij} \hat{\mathbf{g}}_k \right]. \tag{F.25}$$

The quantity in brackets on the RHS is referred to as the covariant derivative of a vector and can be written a bit more compactly as

$$\frac{\partial}{\partial x^j} \left(V^i \hat{\mathbf{g}}_i \right) = \left(\frac{\partial V^k}{\partial x^j} + V^i \Gamma^k_{\ ij} \right) \hat{\mathbf{g}}_k, \tag{F.26}$$

where the Christoffel symbol can always be obtained from Equation F.24. If the basis vectors are constants, $\Gamma^k_{\ ij} = 0$, and the covariant derivative simplifies to

$$\frac{\partial}{\partial x^j} \left(V^i \hat{\mathbf{g}}_i \right) = \left(\frac{\partial V^k}{\partial x^j} \right) \hat{\mathbf{g}}_k, \tag{F.27}$$

as you would expect.

The curl operation can be handled in a similar manner. The curl of the vector field $\overline{V}$ is

$$\overline{\nabla} \times \overline{V} = \hat{\mathbf{g}}^j \frac{\partial}{\partial x^j} \times \left(V^i \hat{\mathbf{g}}_i \right), \tag{F.28}$$

which, written in terms of the covariant derivative, is

$$\overline{\nabla} \times \overline{V} = \left(\frac{\partial V^k}{\partial x^j} + V^i \Gamma^k_{\ ij} \right) \left(\hat{\mathbf{g}}_j \times \hat{\mathbf{g}}_k \right). \tag{F.29}$$

This Page Intentionally Left Blank

G

CALCULUS OF VARIATIONS

The development of the calculus of variations begins with an integral in the following form:

$$I = \int_{x_1}^{x_2} dx \, f[x, y(x), dy/dx]. \tag{G.1}$$

The function in the integrand depends upon the independent variable x and on the unknown function $y(x)$ and its derivative dy/dx. Both $y(x)$ and its derivative are taken to be functions of the independent variable x. The problem is to determine the function $y(x)$ so as to minimize the value of the integral. An example of such an integral is.

$$I = \int_{x_1}^{x_2} dx \left[1 + \left(\frac{dy}{dx}\right)^2\right]^{1/2}. \tag{G.2}$$

This form looks like that of an integral equation because the unknown function is inside an integral operation. But it is unlike an integral equation in that the value of the integral is not known and also because the minimization of I does not uniquely determine $y(x)$ unless a set of boundary conditions like

$$y(x_1) = y_1 \tag{G.3}$$
$$y(x_2) = y_2 \tag{G.4}$$

is specified. Here y_1 and y_2 are known constants.

The minimization process is accomplished by introducing a parameter α so that $I \to I(\alpha)$ and requiring $dI/d\alpha = 0$ at $\alpha = 0$. The α parameter is introduced by letting

$$y(x) \to y(x, \alpha) = y(x) + \alpha\eta(x) \tag{G.5}$$

where $y(x)$ is the function that minimizes $I(\alpha)$ with $\alpha = 0$. The idea is that adding any function $\eta(x)$ with $\alpha \neq 0$ causes $I(\alpha)$ to be larger than $I(\alpha = 0)$.

661

The function $y(x, \alpha)$ and its derivative are then introduced into the integrand of Equation G.1.

$$y(x) \quad \rightarrow \quad y(x, \alpha) \tag{G.6}$$

$$\frac{dy}{dx} \quad \rightarrow \quad \frac{\partial y(x, \alpha)}{\partial x.} \tag{G.7}$$

The boundary conditions

$$y(x = x_1, \alpha) \quad = \quad y_1 \tag{G.8}$$
$$y(x = x_2, \alpha) \quad = \quad y_2 \tag{G.9}$$

require

$$\eta(x_1) = \eta(x_2) = 0. \tag{G.10}$$

A shorthand notation

$$\frac{\partial y(x, \alpha)}{\partial x} = y_x(x, \alpha) \tag{G.11}$$

will be used to simplify the equations.

The problem now becomes one of minimizing $I(\alpha)$ with respect to α where

$$I(\alpha) = \int_{x_1}^{x_2} dx \, f[x.y(x, \alpha), y_x(x, \alpha)] \tag{G.12}$$

under the conditions

$$y(x_1, \alpha) \quad = \quad y_1 \tag{G.13}$$
$$y(x_2, \alpha) \quad = \quad y_2 \tag{G.14}$$
$$\eta(x_1) \quad = \quad 0 \tag{G.15}$$
$$\eta(x_1) \quad = \quad 0. \tag{G.16}$$

To accomplish this take $d/d\alpha$ of the above expression and set it equal to zero at $\alpha = 0$.

$$0 = \left\{ \frac{d}{d\alpha} \int_{x_1}^{x_2} dx \, f[x.y(x, \alpha), y_x(x, \alpha)] \right\}_{\alpha=0} \tag{G.17}$$

or

$$0 = \left\{ \int_{x_1}^{x_2} dx \left[\frac{\partial f}{\partial y} \frac{\partial y(x, \alpha)}{\partial \alpha} + \frac{\partial f}{\partial y_x} \frac{\partial y_x(x, \alpha)}{\partial \alpha} \right] \right\}_{\alpha=0}. \tag{G.18}$$

Now

$$y(x, \alpha) = y(x) + \alpha\eta(x) \tag{G.19}$$

so

$$\frac{\partial y(x, \alpha)}{\partial \alpha} = \eta(x) \tag{G.20}$$

and

$$y_x(x, \alpha) = \frac{\partial}{\partial x} [y(x) + \alpha \eta(x)] \tag{G.21}$$

$$y_x(x, \alpha) = \frac{dy(x)}{dx} + \alpha \frac{d\eta}{dx} \tag{G.22}$$

so that

$$\frac{\partial y_x(x, \alpha)}{\partial \alpha} = \frac{d\eta}{dx}. \tag{G.23}$$

Notice that these partial derivatives of $y(x, \alpha)$ and $y_x(x, \alpha)$ with respect to α are no longer a function of α.

Substituting these values for the partial derivatives of $y(x, \alpha)$ and $y_x(x, \alpha)$ into Equation G.18 gives,

$$0 = \left\{ \int_{x_1}^{x_2} dx \left[\frac{\partial f}{\partial y} \eta(x) + \frac{\partial f}{\partial y_x} \frac{d\eta}{dx} \right] \right\}_{\alpha=0}. \tag{G.24}$$

The second integral

$$\int_{x_1}^{x_2} dx \frac{\partial f}{\partial y_x} \frac{d\eta}{dx}. \tag{G.25}$$

is attacked by an integration by parts with

$$du = \frac{d\eta}{dx} dx \tag{G.26}$$

and

$$v = \frac{\partial f}{\partial y_x} \tag{G.27}$$

so that

$$\int_{x_1}^{x_2} dx \frac{\partial f}{\partial y_x} \frac{d\eta}{dx} = \eta(x) \frac{\partial f}{\partial y_x} \Big|_{x_1}^{x_2} - \int_{x_1}^{x_2} dx \, \eta(x) \frac{d}{dx} \frac{\partial f}{\partial y_x}. \tag{G.28}$$

The first term on the right hand side of this equation vanishes because $\eta(x)$ is zero at x_1 and x_2.

The minimization condition then becomes

$$0 = \left\{ \int_{x_1}^{x_2} dx \, \eta(x) \left[\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y_x} \right] \right\}_{\alpha=0}. \tag{G.29}$$

Since this must be true for any $\eta(x)$ when $\alpha = 0$, the condition for minimizing $I(\alpha)$ becomes

$$0 = \left[\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y_x} \right]_{\alpha=0} \tag{G.30}$$

In this expression

$$f = f[x, y(x, \alpha), y_x(x, \alpha)] \tag{G.31}$$

where
$$y = y(x, \alpha) = y(x) + \alpha \eta(x) \tag{G.32}$$

and the function $y(x)$ is the function we seek, the function that minimizes the original integral. In the limit $\alpha = 0$ the above condition for minimization of the original integral becomes

$$0 = \frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y_x} \tag{G.33}$$

where
$$y = y(x) \tag{G.34}$$

and

$$y_x = \frac{dy(x)}{dx}. \tag{G.35}$$

Equation G.33 is solved for $y(x)$ given the boundary conditions

$$y(x_1) = y_1 \tag{G.36}$$
$$y(x_2) = y_2. \tag{G.37}$$

This is known as Eulers Equation. It can be written in equivalent form as

$$0 = \frac{\partial f}{\partial x} - \frac{d}{dx} \left(f - y_x \frac{\partial f}{\partial y_x} \right). \tag{G.38}$$

This equivalence can be seen by expanding the total derivative with respect to x.

Equation G.33 is useful if there is no explicit y-dependence in the integrand because it reduces to

$$0 = \frac{d}{dx} \frac{\partial f}{\partial y_x} \tag{G.39}$$

Equation G.38 is useful if f has no explicit x-dependence because it reduces to

$$0 = \frac{d}{dx} \left(f - y_x \frac{\partial f}{\partial y_x} \right). \tag{G.40}$$

ERRATA LIST

p. 10   Change Equation 1.48 to read:

$$|\overline{\mathbf{C}}| = |\mathbf{A}|\,|\mathbf{B}|\sin\theta, \qquad\qquad (1.48)$$

p. 16

  (i) Exercise 2, first equation, change c, j to i, j:

$$M_{i,j} = ij^2 \quad \text{for} \quad i, j = 1, 2, 3,$$

  (ii) Exercise 4, first line, change "row matrix" to "column matrix".

  (iii) Exercise 4, first equation, change equation to read:

$$\left[[D][V]\right]^{\dagger} = [V]^{\dagger}[D]$$

p. 49   Change Equation 3.14 to read:

$$\overline{\mathbf{r}} = (\overline{\mathbf{r}}\cdot\hat{\mathbf{e}}_r)\hat{\mathbf{e}}_r + (\overline{\mathbf{r}}\cdot\hat{\mathbf{e}}_\theta)\hat{\mathbf{e}}_\theta + (\overline{\mathbf{r}}\cdot\hat{\mathbf{e}}_\phi)\hat{\mathbf{e}}_\phi. \qquad\qquad (3.14)$$

p. 64   Exercise 20, change equation to read:

$$\overline{\mathbf{B}} = B_o\hat{\mathbf{q}}_z.$$

p. 65   Exercise 20, change equation to read:

$$\overline{\mathbf{B}} = \frac{B_o\rho}{\rho_o}\hat{\mathbf{q}}_\phi,$$

p. 75 Change Equation 4.45 to read:

$$\hat{\mathbf{e}}_i' = a_{ij}\hat{\mathbf{e}}_j. \tag{4.45}$$

p. 133 Exercise 25, first line, change $\rho_c(x, y, y)$ to $\rho_c(x, y, z)$.

p. 152 In all equations on the page, change $\underline{c}_n$ to c_n, i.e.,

$$f(x) = \sum_{n=0}^{\infty} c_n(x - x_o)^n = c_o + c_1(x - x_o) + c_2(x - x_o)^2 + \cdots. \tag{6.96}$$

$$c_n = \frac{1}{n!}\frac{d^n f(x)}{dx^n}\bigg|_{x=x_o} \tag{6.97}$$

$$\lim_{n\to\infty}\left|\frac{c_n(x - x_o)^n}{c_{n-1}(x - x_o)^{n-1}}\right| < 1. \tag{6.98}$$

$$f(x) = \sum_{n=-\infty}^{+\infty} c_n(x - x_o)^n \tag{6.99}$$

$$= \cdots + \frac{c_{-2}}{(x - x_o)^2} + \frac{c_{-1}}{(x - x_o)} + c_o + c_1(x - x_o) + c_2(x - x_o)^2 + \cdots.$$

p. 162 Paragraph below Equation 6.135,

(i) third line, change C_1 to C_2 and C_2 to C_1.

(ii) seventh line, change C_2 to C_1.

p. 215 Exercise 49. (b), change $y = 1/2$ to $u = 1/2$.

p. 216 Exercise 55, change the sentences beginning in the second line of the paragraph to the end of the paragraph to read:
Integrate to obtain $\underline{z} = \underline{z}(\underline{w})$.

p. 241 Change Equation 7.90 to read:

$$f_k = \sum_{n=n_o}^{2N+n_o-1} \underline{c}_n e^{i\omega_n t_k} \tag{7.90}$$

p. 242 Change the first line of Equation 7.92 to read:

$$f_k = \sum_{n=-N}^{N-1} \underline{c}_n e^{i\omega_n t_k}$$

p. 300 Exercise 21:

(i) Change the last sentence of the first paragraph to read:
"Show that the conditions for closure with zero contribution are different for a real, definite integral than they are for a Fourier inversion."

(ii) Change the first equation to read:

$$F(\omega) = \frac{\omega^2}{\omega^3 + i}$$

(iii) Change the equation under (b) to read:

$$f(t) = \int_{-\infty}^{\infty} dx \, e^{i\omega t} F(\omega)$$

p. 316 Figure 9.14, change the label on the abscissa from "imag" to "real".

p. 318 Figure 9.16, change the label on the abscissa from "imag" to "real".

p. 354 Change the part of the sentence below Equation 10.104 to read:

"... where the c_n are unknown coefficents, and s is a fixed number, not necessarily an integer or pure real. Often, however, it is a positive or negative integer."

p. 379 Figure 10.16, change $-K$ x(t) to read $-K_o$ x(t).

p. 383 Title of Figure 10.21, change "Principal" to "Principle".

p. 408 Exercise 22 (a), last line, change $x = x_o$ to $y = y_o$.

p. 421 Exercise 53, first line below the equation, change $a(x), b(x)$ to $p(x), q(x)$.

p. 422 Exercise 54 (d), first line, change $d(t)$ to $d(x)$.

p. 428 Change Equation 11.20 to read:

$$\sum_m \sum_n \left\{ \begin{array}{l} A_{mn}e^{+x\sqrt{c_{nx}}}e^{+y\sqrt{c_{my}}}e^{+z\sqrt{-c_{nx}-c_{my}}} \\ +B_{mn}e^{-x\sqrt{c_{nx}}}e^{+y\sqrt{c_{my}}}e^{+z\sqrt{-c_{nx}-c_{my}}} \\ +C_{mn}e^{+x\sqrt{c_{nx}}}e^{-y\sqrt{c_{my}}}e^{+z\sqrt{-c_{nx}-c_{my}}} \\ +D_{mn}e^{-x\sqrt{c_{nx}}}e^{-y\sqrt{c_{my}}}e^{+z\sqrt{-c_{nx}-c_{my}}} \\ +E_{mn}e^{+x\sqrt{c_{nx}}}e^{+y\sqrt{c_{my}}}e^{-z\sqrt{-c_{nx}-c_{my}}} \\ +F_{mn}e^{-x\sqrt{c_{nx}}}e^{+y\sqrt{c_{my}}}e^{-z\sqrt{-c_{nx}-c_{my}}} \\ +G_{mn}e^{+x\sqrt{c_{nx}}}e^{-y\sqrt{c_{my}}}e^{-z\sqrt{-c_{nx}-c_{my}}} \\ +H_{mn}e^{-x\sqrt{c_{nx}}}e^{-y\sqrt{c_{my}}}e^{-z\sqrt{-c_{nx}-c_{my}}} \end{array} \right\} . \qquad (11.20)$$

p. 448 Split the last sentence of the paragraph ***Orthogonality Relations*** into two sentences:
$$\ldots \text{eigenfunction problem.}$$
A Hermitian operator can then be identified and the orthogonality properties of the eigenfunctions determined.

p. 449 Change Equation 11.111 to read:

$$\left[J_\nu(\alpha_{\nu n}\rho/r_o)\rho \frac{dJ_\nu(\alpha_{\nu m}\rho/r_o)}{d\rho} \right]\Bigg|_\Omega = 0. \tag{11.111}$$

p. 453 Change Equation 11.136 to read:

$$\Phi(\rho,\theta,z) = \sum_\nu \sum_k \begin{Bmatrix} \cos(\nu\theta) \\ \sin(\nu\theta) \end{Bmatrix} \begin{Bmatrix} \cos(kz) \\ \sin(kz) \end{Bmatrix} \begin{Bmatrix} I_\nu(k\rho) \\ K_\nu(k\rho) \end{Bmatrix} \tag{11.136}$$

p. 462 Change Equation 11.176 to read:

$$\dot{A}_\ell(x) = \cdots + a_{n\gg 1}\left\{ x^{n+1} + x^{n+2} + x^{n+4} + \cdots \right\} \tag{11.176}$$

p. 463

(i) Line above Equation 11.180, change $B_\ell(x)$ to $\dot{B}_\ell(x)$.

(ii) Line above Equation 11.181, change $A_0(x)$ to $\dot{A}_0(x)$.

p. 464 Third paragraph, change "six" to "five".

p. 473 Figure 11.21, last line of equation, change $+V_0$ to $-V_0$.

p. 474 Equation 11.237: replace closing brace by opening brace:

$$\Phi(r,\theta,\phi) = \sum_{\ell=0}^\infty \sum_{m=-\ell}^{+\ell} \left\{ \begin{matrix} r^\ell \\ r^{-\ell-1} \end{matrix} \right\{ \underline{Y}_{\ell m}(\theta,\phi). \tag{11.237}$$

p. 482 Exercise 16(a), change Equation to read:

$$z(\rho) = \int_o^{r_o} d\xi\, F(\xi)g(\rho \mid \xi).$$

p. 484 Exercise 24, change the second and third line below the first equation to read:
"…the differential equation for radially dependent component becomes"

p. 489 Exercise 34: insert 0 on the right-hand side of Laplace's equation in two dimensions:

$$\nabla^2 \Psi(q_1, q_2) = \frac{1}{h_1 h_2} \left[\frac{\partial}{\partial q_1} \frac{h_2}{h_1} \frac{\partial}{\partial q_1} + \frac{\partial}{\partial q_2} \frac{h_2}{h_1} \frac{\partial}{\partial q_2} \right] \Psi(q_1, q_2) = 0,$$

p. 507 Exercise 7, change equation to read:

$$\psi(x) = x - \int_0^x dt \, (t - x)\psi(t).$$

p. 508 Exercise 12.i., change equation to read:

$$\psi(x) = \lambda \int_{-1}^1 dt \, (x - t)^2 \psi(t).$$

p. 537 First line of the paragraph above Equation 13.45, change sentence to read: "We should also look carefully at the little circular part. . . "

p. 542

(i) Second line from top of page, change term to read $d\underline{z} = e^{i2\pi} dr$.

(ii) Change Equation 13.57 to read:

$$\oint_{C_o} = 2\pi = \int_1 + \int_3 = 2I. \tag{13.57}$$

p. 547 Change Equation 13.72 to read:

$$\frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} = \frac{\partial \underline{w}}{\partial x} = \frac{\partial \underline{w}}{\partial \underline{z}} = 0 \quad \text{at the saddle points.} \tag{13.72}$$

p. 581 Change Equation 14.100 to read:

$$= t^i_j \frac{\partial}{\partial x_j}, \tag{14.100}$$

p. 593 Exercise 3, change equation to read:

$$\overline{\overline{\mathbf{M}}} = M_{ij} \hat{\mathbf{g}}^i \hat{\mathbf{g}}^j = M^{ij} \hat{\mathbf{g}}_i \hat{\mathbf{g}}_j.$$

p. 611 In the first line of the paragraph above Table 15.16, change "are shown" to "is shown".

p. 614 In the fourth line from the top of the second paragraph, change "dimension" to "dimensions".

p. 619   Table 15.25, row $C_4^{[2]}$, column g_4, change "-1" to "$-i$".

p. 623   Change the sentence beginning in the 2nd line of the 2nd paragraph to read:

"The net result is that there is one $D_3^{[1]}$ representation, no $D_3^{[2]}$ representation, and one $D_3^{[3]}$ representation in the block diagonalization."

p. 625   Equation 15.62, change element (1,3) and (3,1) from "1" to "-1".

BIBLIOGRAPHY

Abramowitz, Milton, and Irene Stegun, *Handbook of Mathematical Function.* Dover Publications, New York, 1964.

Arfken, George, *Mathematical Methods of Physics.* Academic Press, San Diego, CA, 1985.

Butkov, Eugene, *Mathematical Physics.* Addison-Wesley, Reading, MA, 1968.

Churchill, Ruel V., and James Ward Brown, *Complex Variables and Applications.* McGraw-Hill, New York, 1990.

Griffel, D.H., *Applied Functional Analysis.* John Wiley and Sons, New York, 1981.

Hamermesh, Morton, *Group Theory and its Applications to Physical Problems.* Addison-Wesley, Reading, MA, 1962.

Hildebrand, F.B., *Methods of Applied Mathematics.* Prentice-Hall, Englewood Cliffs, NJ, 1958.

Jackson, J.D, *Classical Electrodynamics.* John Wiley and Sons, New York, 1962.

Kaplan, Wilfred, *Advanced Calculus.* Addison-Wesley, Reading, MA, 1973.

Mathews, Jon, and R.L. Walker, *Mathematical Methods of Physics.* Addison-Wesley, Redwood City, CA, 1970.

Misner, Charles W., Kip S. Thorne, and Jon Archibald Wheeler, *Gravitation.* W.H. Freeman, San Francisco, CA, 1973.

Sokolnikoff, Ivan, and R.M. Redheffer, *Mathematics of Physics and Modern Engineering.* McGraw-Hill, New York, 1966.

Taylor, Edwin F., and Jon Archibald Wheeler, *Spacetime Physics.* W.H. Freeman, San Francisco, CA, 1963.

Thomas, George B. Jr., and Ross L. Finney, *Calculus and Analytic Geometry.* Addison-Wesley, Reading, MA, 1951.

Tolstov, Georgi P., *Fourier Series,* translated by Richard A. Silverman. Dover Publications. New York, 1962.

Tung, Wu-Ki, *Group Theory in Physics.* World Scientific Publishing, Philadelphia, PA, 1985.

Watson, G.N., *A Treatise on the Theory of Bessel Functions.* Macmillian, New York, 1944.

Wigner, Eugene Paul, *Group Theory and its Applicatons to the Quantum Mechanics of Atomic Spectra.* Academic Press, New York, 1959.

This Page Intentionally Left Blank

INDEX

This Page Intentionally Left Blank

This Page Intentionally Left Blank

This Page Intentionally Left Blank

This Page Intentionally Left Blank

This Page Intentionally Left Blank

This Page Intentionally Left Blank